MACROEVOLUTIONARY PATTERNS AND PROCESSES OF DIVERSIFICATION IN SEDGES (CYPERACEAE), WITH EMPHASIS ON ELEOCHARIS

By

CODY HINCHLIFF

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

DECEMBER 2011

© Copyright by CODY HINCHLIFF, 2011 All Rights Reserved
To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of CODY HINCHLIFF find it satisfactory and recommend that it be accepted.

_______________________________________
Eric H. Roalson, Ph.D., Chair

_______________________________________
Luke J. Harmon, Ph.D.

_______________________________________
Jeremiah W. Busch, Ph.D.
ACKNOWLEDGEMENT

I would first like to thank my family, Peter, Gaye, and Caroline Hinchlif, for encouraging me throughout my life to pursue a career in botany. Of equal importance have been my advisor Eric Roalson, my undergraduate mentors Frederica Bowcutt and Jack Longino, and my peers and friends on the Palouse and elsewhere, Chad Brock, Tim Carey, John Clark, Luke Harmon, Sarah Jacobs, Matt King, K. Marlowe, Mare Nazaire, James Riser, John Schenk, David Tank, Simon Uribe-Convers, and Wayne Warrington, to name a few, who have supported me in countless ways for the last six years and made it possible for this research to be completed. From the Smithsonian, Karen Redden, Ken Wurdack, and Mark Strong gave me the rare opportunity to experience the biodiversity of Guyana and include it in this work. Our collecting efforts there would not have been possible without the support of Claudius Perry, Paul Benjamin, Dave Singh, Elford Liverpool, and the Village of Isseneru. The friendly and helpful staff of the Australian Northern Territory and Sydney herbaria, Ian Cowie, Bob Harwood, Dr. Dale Dixon, and Dr. Leonie Stanberg were instrumental in supporting Australian field work. In Bolivia, Ariel Liuiy Aguilar wins the award for best field collaborator ever, thank you Ariel for all your help, trust, and friendship. When I think of South America, your smiling face comes to mind first. Thanks also to Dr. Stephan Beck, Oscar Plata, Rosa Isela Meneces, and the staff of the Herbario Nacional for their invaluable and ongoing help and support with Bolivian collections. Steve Orzell, Edwin Bridges, and Eric Menges were instrumental in finding specimens of Websteria in Florida, and I had a fantastic time seeing their state through their eyes. Craig Thurman of Arab, Alabama graciously provided the otherwise impossible to find specimens of Egleria. This work has been supported by NSF grant 0515975 to E. H. Roalson. The python script used to calculate $I^S$ scores in Chapter 3 is available by request from C. E. Hinchlif.
MACROEVOLUTIONARY PATTERNS AND PROCESSES OF DIVERSIFICATION IN
SEDGES (CYPERACEAE), WITH EMPHASIS ON ELEOCHARIS

Abstract

by Cody Hinchliff, Ph.D.
Washington State University
December 2011

Chair: Eric H. Roalson

This dissertation includes three chapters, the first two of which focus on the genus
Eleocharis—an ecologically and economically important group of sedges containing over 250
species, with centers of diversity in seasonally wet tropical to subtropical regions. In Chapter 1,
we examine morphological evolution of stems in Eleocharis subgenus Limnochloa. In Eleocharis,
the stems serve as the primary photosynthetic organs and as such their structure is presumed to
affect photosynthetic efficacy. We identify a complex history of stem shape evolution
characterized by a high degree of homoplasy and rapid rates of change, using stochastic mapping
and Markov 1-rate models to infer evolutionary changes. Our data also suggest that changes in
stem shape and anatomy may be associated with speciation (Pagel’s $\kappa = 0.3503$, $p = 0.04579$).
The second chapter focuses on intrageneric relationships in Cyperaceae, with emphasis on
Eleocharis. We elucidate the relationships between Eleocharis and two monotypic genera—
Egleria and Websteria—which we show to be included within Eleocharis. In addition, we clearly
document a sister relationship between Eleocharis and the Abildgaardieae. A third questionably
segregate genus, Chillania, could not be included in the molecular dataset, but its morphology
suggests it should be included in Eleocharis. Supporting nomenclature is provided to place
Chillania, Egleria, and Websteria in Eleocharis. The final chapter focuses on infrafamilial
relationships within Cyperaceae, as well as the exploration of novel phylogenetic methods to
reconstruct histories using sparse supermatrices mined from public nucleotide sequence
repositories (in this case GenBank). We show that highly incomplete datasets can yield very good
estimates of phylogeny, and we offer suggestions for ways to improve data decisiveness and
phylogenetic utility of datasets by filtering taxa and genes. We present the best-resolved
phylogenetic trees of the Cyperaceae that have been published to date, and we discuss their
significance for sedge classification.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ........................................................................................................ iii

**ABSTRACT** ........................................................................................................................ iv-v

**LIST OF TABLES** ................................................................................................................ ix

**LIST OF FIGURES** ............................................................................................................... x

**GENERAL INTRODUCTION** ................................................................................................. 1

**CHAPTER 1: STEM ARCHITECTURE IN *ELEOCHARIS* SUBGENUS *LIMNOCHLOA* (CYPERACEAE): EVIDENCE OF DYNAMIC MORPHOLOGICAL EVOLUTION IN A GROUP OF PANTROPICAL SEDGES**

1. **INTRODUCTION** .............................................................................................................. 2

2. **MATERIALS AND METHODS** .......................................................................................... 6
   Taxon sampling......................................................................................................................... 6
   DNA extraction, PCR, and sequencing..................................................................................... 7
   Phylogenetic analyses ............................................................................................................ 7
   Character reconstructions ..................................................................................................... 8
   Modeling patterns of diversification..................................................................................... 10

3. **RESULTS** ........................................................................................................................ 11

4. **DISCUSSION** .................................................................................................................. 15
   Character reconstructions ..................................................................................................... 16
   Evolutionary implications of stem shape .............................................................................. 19
   Developmental basis of stem structural evolution ............................................................... 22
   Concluding remarks ............................................................................................................. 23

5. **LITERATURE CITED** ....................................................................................................... 24

1. INTRODUCTION ........................................................................................................................................... 38

2. MATERIALS AND METHODS .................................................................................................................. 41
   Plant samples ................................................................................................................................................... 41
   DNA sequencing ............................................................................................................................................. 42
   Phylogenetic analyses ................................................................................................................................. 43

3. RESULTS .................................................................................................................................................... 44

4. DISCUSSION ............................................................................................................................................. 45
   General relationships in the Cyperaceae ................................................................................................. 45
   Relationships within the Eleocharideae ................................................................................................. 46
   Implications for classification ..................................................................................................................... 49

5. LITERATURE CITED ................................................................................................................................. 53

CHAPTER 3: REFINING A SUPERMATRIX OF SEDGES: PRACTICAL METHODOLOGY AND CONCLUSIVE RESULTS FROM A LARGE COMBINED ANALYSIS OF PUBLICLY AVAILABLE SEQUENCE DATA

1. INTRODUCTION ........................................................................................................................................... 68

2. METHODS .................................................................................................................................................. 69
   Data collection and validation ....................................................................................................................... 69
   Data filtering and heuristic searches .......................................................................................................... 72

3. RESULTS .................................................................................................................................................... 73

4. DISCUSSION ............................................................................................................................................. 76
   Implications for large, combined analyses ................................................................................................. 76
   Sequence data sampling strategies .............................................................................................................. 79
   Systematics of Cyperaceae ........................................................................................................................ 81
5. LITERATURE CITED .................................................................................................................. 84

APPENDIX

A. LIST OF VOUCHERED DNA SAMPLES FOR CHAPTER 1 .................................................. 98
B. LIST OF VOUCHERED DNA SAMPLES FOR CHAPTER 2 ................................................... 102
C. FORMULAE FOR SUMMARY STATISTICS USED IN FIGURE 11 ................................. 109
LIST OF TABLES

1. Mean transition frequencies among stem states as inferred by stochastic mapping .......... 15
2. Descriptive statistics for alignments and trees resulting from each of four combined approaches to data filtering ................................................................................................................. 75
LIST OF FIGURES

1. Figure 1; Stem structural variation in *Eleocharis* subgenus *Limnochloa* ........................................31
2. Figure 2; Maximum likelihood phylogeny of *Eleocharis* subgenus *Limnochloa* .......................32
3. Figure 3; Character reconstructions for stem states in *Eleocharis* subgenus *Limnochloa* ........33
4. Figure 4; Transition probabilities among stem character states in *Eleocharis* subgenus *Limnochloa* generated by stochastic mapping .................................................................35
5. Figure 5; Likelihood surface for best-fit value of Pagel’s $\kappa$ associated with stem state transitions in *Eleocharis* subgenus *Limnochloa* ........................................................................................................36
6. Figure 6; Effect of applying the best-fit value of Pagel’s $\kappa$ to an ultrametric phylogeny of *Eleocharis* subgenus *Limnochloa* ........................................................................................................37
7. Figure 7; Phylogeny of Cyperaceae based on *ndhF* and *psbB-psbH* data, part I .................60
8. Figure 8; Phylogeny of Cyperaceae based on *ndhF* and *psbB-psbH* data, part II ............62
9. Figure 9; Morphology of *Egleria*, *Eleocharis*, and *Websteria* ......................................................64
10. Figure 10; Morphology of *Eleocharis* and *Chillania* ..........................................................66
11. Figure 11; Heat map of nucleotide sequence sampling in Cyperaceae in GenBank release 185.................................................................................................................................................89
12. Majority-rule consensus trees resulting from parallel 300-replicate ML bootstrap searches performed on each of the alignments summarized in Table 2 ........................................91
13. Majority rule consensus tree resulting from a 300-replicate ML search performed on the alignment corresponding to Table 1, row 4, part I .................................................................93
14. Majority rule consensus tree resulting from a 300-replicate ML search performed on the alignment corresponding to Table 1, row 4, part II ........................................................................95
15. Majority rule consensus tree resulting from a 300-replicate ML search performed on the alignment corresponding to Table 1, row 4, part III .................................................................96
16. Majority rule consensus tree resulting from a 300-replicate ML search performed on the alignment corresponding to Table 1, row 4, part IV ............................................................ 97
Dedication

This work is dedicated to David Michael McDowell and Jacob Theodore Peterson. Your love and companionship made the last six years of my life the best ones yet, and I could not have done this without your support.
GENERAL INTRODUCTION

This dissertation integrates three manuscripts that were completed through the course of my studies at Washington State University between the years of 2006 and 2012. These manuscripts detail my studies in the genus *Eleocharis* (Cyperaceae), as well as broader-scale studies exploring patterns of diversification in the family Cyperaceae itself. Chapter 1 is a phylogenetically informed morphological study of an important, diverse tropical clade of *Eleocharis*, and was published in in 2009 in Systematic Botany 96, pp. 1487-1499. It follows the standard formatting conventions for that journal. Chapter 1 was jointly authored with Eric Roalson, who provided methodological training, general feedback, and editorial suggestions. Chapter 2 was published in 2010 in Taxon 59, pp. 709-719, and follows the standard formatting guidelines for that journal. It addresses intrageneric relationships in the Cyperaceae, with special emphasis on two genera previously classified as segregates of *Eleocharis*, which we showed to be nested within that genus. The according nomenclatural changes are presented. Chapter 2 was jointly authored with Ariel Lliully Aguilar, of Bolivia, who provided field expertise and conducted collecting trips, Timothy Carey, who contributed lab time and sequence data, and Eric Roalson, who provided his expertise with Cyperaceae phylogeny and editorial suggestions. Chapter 3 expands the focus even further, and presents a very broadly sampled sedge phylogeny inferred from aggregated sequence data from GenBank. This chapter addresses family-level relationships in the Cyperaceae and presents resolutions to some long-standing questions of phylogenetic placement, but of greater interest to the systematics community may be the methods we used to develop our datasets and improve the resolution of the resulting trees, despite having very incomplete coverage within our character matrices. This final chapter will be submitted to Systematic Biology, and we expect it to be published in 2012. It follows the formatting guidelines for that journal. Chapter 3 was jointly authored with Eric Roalson, who provided theoretical support and editorial suggestions.
CHAPTER 1: STEM ARCHITECTURE IN *ELEOCHARIS* SUBGENUS *LIMNOCHLOA* (CYPERACEAE): EVIDENCE OF DYNAMIC MORPHOLOGICAL EVOLUTION IN A GROUP OF PANTROPICAL SEDGES

INTRODUCTION

The cosmopolitan genus *Eleocharis* R.Br. (Cyperaceae) contains about 250 species and is common in most regions of the world, from sea level to above 5,000 m elevation. Growth forms include herbaceous annuals and perennials, which can be caespitose, mat-forming, or rhizomatous, and grow from heights of less than one centimeter to greater than three meters. They are generally found in wet places, often where there is strong seasonal variation in water level. All species of *Eleocharis* lack leaf blades entirely, having instead reduced leaf sheaths at the base of their photosynthetic stems. The stems function as the primary carbon-fixing organs of the plant (Baksh and Richards, 2006). Phylogenetic studies have suggested that the transition from C$_3$ to C$_4$ has evolved at least twice within the genus (Roalson and Friar, 2000), and its species show diverse photosynthetic adaptations, including C$_3$, C$_4$, and intermediate pathways (Roalson, 2007). The photosynthetic pathway plasticity seen in *Eleocharis vivipara* Link is particularly novel. This species can switch between C$_3$ and C$_4$ depending on its growing conditions (Ueno, 1988; Murphy et al., 2007), and under some circumstances, may include other carbon concentrating mechanisms such as a bicarbonate pump (Murphy et al., 2007).

Infrageneric relationships in *Eleocharis* are not well understood. Because of high levels of homoplasy and a presumed lack of phylogenetically informative morphological characters across the genus (Gonzalez-Elizondo and Tena-Flores, 2000), the determination of evolutionary relationships has been problematic. Svenson (1929, 1934, 1937, 1939) published a detailed series of monographs on *Eleocharis*, describing numerous species and delimiting many infrageneric taxa in the process, using morphological data. Although some of Svenson’s taxa were later identified as poly- or paraphyletic following molecular phylogenetic studies, a number of them represent monophyletic groups, including
the subgenus *Limnochloa* B. Beauv ex T. Lestib. (referred to by Svenson as series *Mutatae*), which is sister to the rest of the genus (Yano et al., 2004; Roalson and Friar, 2000).

Subgenus *Limnochloa*, which constitutes about 1/10th of all *Eleocharis* species (about 30 out of about 300 species) (Govaerts and Simpson, 2007), is circumscribed by a combination of traits including generally cylindrical spikes that are rarely wider than the culms, spikelet scales with 15 or more often prominent longitudinal veins, and relatively large achenes that are usually markedly sculptured (González-Elizondo and Peterson, 1997). Most species of subgenus *Limnochloa* are emergent aquatic rhizomatous perennials of the lowland tropics, although a number are found in the temperate zones as far north as 49° latitude (*Eleocharis equisetoides* (Elliot) Torr., *E. robbinsii* Oakes, and *E. quadrangulata* (Michx.) Roemer: Ontario, Canada) and as far south as 29° latitude (*E. kleinii* Barros: Santa Catarina, Brazil). Subgenus *Limnochloa* contains many species of vital ecological importance, which provide a significant proportion of the food and habitat resources required by numerous animal species. This ecological role is particularly apparent in the coastal floodplains of Australia’s Northern Territory (Finlayson, 2005; Bayliss and Yeomans, 1990) and other lowland tropical wetlands where species of subgenus *Limnochloa* dominate the ecosystem (C. Hinchliff, personal obs.). The subgenus also contains the most prevalent agricultural crop species in the genus, *E. dulcis* Trin ex. Henschel, the Chinese water chestnut.

The genus *Eleocharis* has only recently become the subject of rigorous phylogenetic analysis, and the bulk of this research remains unpublished (unpublished data). As such, species-level relationships within the genus are not well understood. Two previous phylogenetic studies of *Eleocharis* exist (Yano et al. 2004; Roalson and Friar 2000), but neither included more than minimal taxon sampling, and left relationships among many species untested. Contributing to the uncertainty surrounding *Eleocharis* are some currently accepted species descriptions that circumscribe extremely morphologically variable and geographically widely dispersed species (e.g., *E. dulcis*, *E. acicularis*, *E. minima*, and others), whose monophyly is questionable. Finally, although new species are being described at a relatively rapid rate (Rosen and Hatch, 2007; Trevisan et al., 2007, Trevisan and Boldrini, 2006; González-Elizondo and
Reznicek, 1996), many yet undescribed species are known (C. E. Hinchliff and E. H. Roalson, unpublished data; K. Wilson, unpublished data) and some species are likely awaiting discovery. As the first phylogenetic study of *Eleocharis* to attempt sorting these species-level relationships, this paper marks an important first step in understanding the evolution of this unusual and biologically significant genus.

*Eleocharis* species show extreme variation in stem architecture. Mature stem diameter varies by nearly two orders of magnitude from less than one half of one millimeter in the narrowest-stemmed, mostly aquatic species to more than two centimeters in the largest, which often grow as emergent aquatics but may also grow terrestrially in close proximity to water or in seasonally inundated habitats. Further, stem geometric shape varies dramatically along a spectrum (i.e. not discretely) from cylindrical to laterally flattened, three-, four-, or five-angled (in *E. yecorensis* Roalson, which has the appearance in cross-section of a five-pointed star), and stem internal architecture (Fig. 1) varies more or less discretely among three major stem types:

(i) Spongy (or non-septate) culms, which are common in subgenus *Limnochloa*, contain extensive networks of incomplete transverse and longitudinal septae that separate numerous small (typically < 1 mm width, regardless of the width of the culm), irregularly shaped, air-filled cavities (note: these are multi-cellular structures that are not analogous to the plants’ cells). Vascular bundles are typically scattered more or less arbitrarily throughout the spongy matrix of septae and air cavities inside the culm.

(ii) Transversely septate culms, which are uncommon both in subgenus *Limnochloa* and in the genus as a whole, have complete transverse septae that separate large air-filled cavities inside the culm. These culms are completely hollow except for the sometimes quite thin (< 0.1 mm thick) transverse septae. Vascular bundles are present only in the culm wall.

(iii) Stems characterized as transversely septate with a central vascular bundle are rare in subgenus *Limnochloa* and absent from the rest of the genus. These culms are similar to transversely septate culms but they bear a single, relatively large vascular bundle that extends the length of the culm.
though its geometric center. This bundle passes through the transverse septae within the otherwise hollow stem.

Of all the lineages within *Eleocharis*, subgenus *Limnochloa* shows the greatest variation in gross stem morphology. The taxa within this subgenus exhibit all the geometric forms and internal anatomical characteristics found within the entire genus. In fact, some *Eleocharis* stem characteristics (such as the five-pointed star lateral profile) occur only in subgenus *Limnochloa* species. This exceptionally large degree of variation in stem morphology across such a short phylogenetic distance seems likely to be due to a rapid rate of change in stem characters within this clade. Despite the large amount of variation in these characters and their close association with the photosynthetic apparatuses of these plants (i.e. their culms), no ecological mechanisms driving the evolution of these characteristics have been identified. This is due primarily to a lack of research on this topic. This study is in fact the first study that concerns the evolution of stem structural characteristics in more than a single species of *Eleocharis*, and one of the only studies to examine the evolution of stem structure in general (but see Hearn, 2006), and as such, at this early stage we are concerned primarily with observing and recording the patterns associated with stem structural change in this lineage. We also offer some conjecture regarding processes potentially related to these patterns, but it is outside the scope of this study to do more than briefly discuss process-related questions.

Subgenus *Limnochloa* is an excellent model for initiating a study into the evolution of stem morphology in herbaceous monocots, because few to no other monocotyledonous taxa show such a pronounced diversity of stem types across a similar phylogenetic distance (although some other genera of Cyperaceae, such as *Fuirena*, *Schoenoplectus*, and *Cyperus* may come close; see Metcalfe (1971)). If evolutionary patterns of stem structural change observed in subgenus *Limnochloa* hold as well in *Eleocharis* as a whole, or in other groups of monocots, our predictive power regarding the potential evolution and adaptive roles of stem characteristics across the monocots would be greatly improved.

Our intention in this study was to address two related questions: (1) what are the phylogenetic relationships among the taxa in this clade, and (2) what are the observed patterns of stem architectural
diversification in *Eleocharis* subgenus *Limnochloa*? We also address a related question of interest: how do the observed patterns of stem evolution relate to the general patterns of diversification in the subgenus? This question is a synthesis of the former two questions, and provides insight into overall patterns of diversification within *Eleocharis* and also additional resolution to our knowledge of patterns of diversification in herbaceous monocots in general.

We used two different approaches to ancestral state reconstruction: (i) stochastic mapping, developed by Huelsenbeck et al. (2003), and (ii) the likelihood-based Markov k-state one rate (Mk1) model (Lewis, 2001) as implemented in Mesquite 2.0b1 (Maddison and Maddison, 2007). Both these models predict evolutionary processes but do so in different ways (Huelsenbeck et al., 2003; Lewis, 2001). To infer phylogenies, we used maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference.

**MATERIALS AND METHODS**

*Taxon sampling*—Thirty-seven taxa representing about 80% of the described specific and subspecific diversity among all of the major lineages of subgenus *Limnochloa* were sampled. Six species from the *E.* subgenera *Zinserlingia* and *Eleocharis*, which were indicated to share a sister relationship with subgenus *Limnochloa* (Hinchliff & Roalson, unpublished data), were used as outgroup taxa in the analyses. Tissue samples collected from wild plants and preserved in silica gel were used whenever possible. For those taxa not collected in silica gel, samples were taken from herbarium specimens. Voucher specimens for all wild-collected plants are deposited at WS (Appendix I). Several species, especially from tropical Africa and South America, could not be included because of the general lack of recent collections of these species in the herbarium record. One of these African species, *E. decoriglumis* Berhaut, shows 3-angled stem morphology that is uncommon within the group. Other 3-angled taxa from subgenus *Limnochloa* occurring in Africa, Australia and the New World tropics were included.
**Morphological measurements**—Stem morphology was evaluated by examining fresh and dried herbarium specimens by eye and using a low power (c.a. 30x) dissecting microscope. The characters used in this study are macromorphological and it was not necessary to section, stain, or examine the material with a more highly powered microscope in order to observe them.

**DNA extraction, PCR, and sequencing**—Total genomic DNA was extracted from all tissue samples using a modified CTAB extraction procedure (Doyle and Doyle, 1987). The polymerase chain reaction (PCR) was used to amplify the nrDNA ITS-1, ITS-2, and 5.8S regions using the ITS-4 and ITS-5 primers (Roalson and Friar, 2004). The chloroplast DNA spacer regions trnC-ycf6 and ycf6-psbM were amplified using the protocols and primers sequences described in Shaw et al. (2005). PCR products were cleaned using the ExoSAP-IT™ procedure (USB Corporation, Cleveland, Ohio, USA). Cycle sequencing reactions were run using the same primers used in the amplification reactions and BigDye™ Terminator 3 (Applied Biosystems, Foster City, California, USA) under a protocol modified for 1/10th reactions. Products of the cycle sequencing reactions were run on an Applied Biosystems 3730 DNA Analyzer at the Washington State University School of Biological Sciences Center for Integrated Biotechnology.

**Phylogenetic analyses**—Sequences were edited in the program Sequencher 4.1.2 (Gene Code Corp., Ann Arbor, Michigan, USA) and manually aligned using the program SeAl 2.0 (Rambaut, 2002). Parsimony analyses were executed in PAUP*4.0b10 (Swofford, 2001) using a heuristic search algorithm with $10^6$ replicate searches. A random addition sequence was used to generate a starting tree at the beginning of each replicate. Indels were treated as missing data. Parsimony bootstrap proportions were calculated in PAUP*4.0b10 from 500 heuristic search bootstrap replicates run to 10,000 generations each.

A heuristic search algorithm was implemented in PAUP*4.0b10 to approximate the maximum likelihood tree. The corrected Akaike Information Criterion (AICc) was used to evaluate the best fit of all possible models to our data (Posada and Buckley, 2004). The selected model corresponded to a submodel of the GTR+Γ substitution model, with a gamma shape parameter value of 0.226233. We implemented a
discrete gamma approximation with six categories in PAUP, and used the best-fit rate matrix values
determined by the AICc tests. The heuristic search consisted of 100 replicates, each of which used a
random addition sequence and had a rearrangement limit set at 20,000. To further optimize the heuristic
search, only 100 trees with a -lnL score lower than 6,500 were saved for branch swapping in each
replicate. No limit was set for trees with a score greater than or equal to 6,500 (an arbitrary value which
was determined in previous analyses to be close to the -lnL score of the best tree). A maximum likelihood
bootstrap analysis was also run in PAUP*4.0b10 using a substitution model identical to the model used in
the search for the best tree. We ran 500 bootstrap replicates, using a heuristic search algorithm with
random addition and a limit of 100,000 rearrangements/replicate.

Posterior probabilities for branches in the tree were assessed using the program MrBayes version
3.1.2 (Huelsenbeck and Ronquist, 2001). A GTR model of DNA evolution with 6 substitution types and a
4-category gamma rate was used for all marker regions. The gamma shape, branch length, proportion of
invariable sites, and substitution rates parameters (shape, brlens, pinvar, and revmat respectively) were
allowed to vary independently for each marker region, while the topology was constrained. Two
simultaneous Bayesian analyses, each using 1 cold chain and 3 heated chains, were run for 4.0 \times 10^7
generations. Trees and parameter values were sampled every 1,000 generations, generating distributions
of 40,000 trees for each marker region. The first 20% of trees from each analysis were discarded as burn-
in—this interval included all generations of the process that occurred before stationarity was reached. All
remaining trees were used to generate a 50% majority rule consensus tree, where the percentage of
occurrence of each branch in the distribution of trees represents the posterior probability (PP) of that
branch existing in the true tree. Branches with a PP \geq 0.95 were considered strongly supported by the data.

**Character reconstructions**—Stochastic mapping, as implemented in the program SIMMAP
(Bollback, 2006), was used to simulate the evolution of stem morphology. SIMMAP uses a stochastic
algorithm to map discrete character states onto a distribution of phylogenetic trees and then summarizes
character history statistics across all individual mappings, thereby incorporating topological and branch
length uncertainty contained in the distribution of trees (Nielsen, 2002; Huelsenbeck et al., 2003). We used the posterior distribution of trees generated in our MrBayes analyses to create stochastic mappings of stem architectural change in SIMMAP. Stem architectural state was assumed to vary among the seven unique combinations of stem geometric shape and internal architecture that are observed in extant taxa (other combinations, such as 3-angled/septate, do not exist). The observed states, which were numbered 0-6, are as follows: (0) round/spongy; (1) round/septate; (2) round/septate with a central vascular bundle; (3) 3-angled-rounded/spongy; (4) 3-angled-acute/spongy; (5) 4-angled/spongy; (6) 5-angled (i.e., star-shaped)/spongy. Each of these states represents a unique combination of one type of gross external stem shape (Fig. 1 I-V) and one type of stem internal anatomy (Fig. 1 A-C). The 3-angled-rounded category included species showing a consistent morphological pattern of rounded stems that became more or less triquetrous toward their apices (e.g., *E. spiralis*). The 3-angled-acute category included species with no rounding of the stem angles from the apex to the base of the stem (e.g., *E. acutangula*).

We generated one character history mapping in SIMMAP for each tree in the posterior distribution, excluding burn-in trees (20%). We used a gamma rate prior, with the shape parameters $\alpha = 7.0$ and $\beta = 2.0$. The reconstruction of character histories was robust with respect to this prior; simulations run under other values of $\alpha$ and $\beta$ generated very similar results. Rates of change among all character states were averaged across all character history reconstructions. Inferred states of internal nodes of the tree were expressed as the proportion of trees in the distribution in which each node was reconstructed to be at each character state. These proportions reflect the uncertainty in the reconstruction of the state of each node, which in turn is dependent on the branch lengths and topologies of all the trees in the distribution. Character states were reconstructed for all internal nodes, though posterior support is low (< 50% PP) for some nodes (Fig. 2). Interpretation of reconstruction proportions for nodes with low support values may not be straightforward.

We also used SIMMAP to summarize the number of transitions among character states on each history in the distribution of simulated histories generated by SIMMAP, and averaged these per-tree
values to yield a mean per-tree number of transitions for each possible combination of all character states in the analysis.

Ricklefs (2007), suggested that because the statistical properties of stochastic mapping and other recent ASR models are not well understood, results should be compared among them. We chose to compare our SIMMAP results to an inferred ASR history generated by Mesquite 2.01 build j28 (Maddison and Maddison, 2006, 2007), under a Markov 1-rate model. The Mk1 method incorporates only a single topology. We used the maximum likelihood topology generated by PAUP* for this purpose.

**Modeling patterns of diversification**—To test for the signature of patterns of diversification other than neutral drift, we used the function fitDiscrete, contained in the phylogenetic diversification analysis package geiger 1.2-13 (Harmon et al., 2007) for the R 2.7.0 statistical framework (R Development Core Team, 2008). fitDiscrete uses ML to estimate the best fit of the observed character data to a topology under several different evolutionary models. Likelihood ratio tests can then be used to determine if some models provide a significantly better fit than others. We chose to test the fit of the data under one null model (the only parameter in this model being the rate of change among character states, see next paragraph for elaboration) and two different model scenarios described by Pagel (1994, 1997). These models added a single parameter each: $\kappa$ and $\delta$, which are each associated with a different pattern of diversification. The parameter $\kappa$ is an exponent applied to the length of each branch in the tree; as $\kappa$ approaches 0, the length of every branch in the tree approaches 1. A high likelihood for a low value of $\kappa$ suggests that the best fit of the character data to the tree is “speciation”, which is to say that most or all character transitions are associated with a single speciation event. When $\kappa$ is close to 1, the branch lengths of the tree are hardly modified at all and there is little relationship between branching events and character transitions. The parameter $\delta$ is an exponent of the depths of all nodes in the tree. When $\delta < 1$, the depths of branching events are lessened, which means that branch lengths toward the base of the tree are lengthened, and evolution is concentrated early in the tree. When $\delta > 1$, the opposite scenario is
created, and evolution is concentrated late in the tree. A \( \delta \) value close to 1 indicates there is little bias in either direction.

We assumed an equal rate of change among character states for all models employed for this analysis. This is not a biologically realistic assumption for this clade, but the high number of states in the rate matrix and the relatively low number of taxa in our analysis preclude using a more robust assumption set regarding rates of change; the likelihood surface for the values of the rate matrix becomes increasingly flatter with the addition of more rates (i.e., more parameters), and the apparent number of transitions that have actually occurred in the history of the clade is so low that some rates (for instance, the transition from state 0 to state 6) simply cannot be directly estimated from the data. The assumption of equal rates should not invalidate the significance of the fit of either of the diversification models to the data (L. Harmon, University of Idaho, pers. comm.).

fitDiscrete requires a topology presented as a chronogram (i.e. with branch length calibrated to relative or absolute time) in order to estimate meaningful rates of diversification. Fossils are lacking for *Eleocharis* subgenus *Limnochloa*, so we could not calibrate our chronogram to absolute time. Instead we used penalized likelihood rate-smoothing as implemented in the program r8s to generate an ultrametric tree (Sanderson, 2002), with the scale of the chronogram set to 1. The rate-smoothing algorithm scales branch lengths to represent relative node ages; absolute age estimates are not generated. Smoothing parameters were derived using cross-validation (data not shown). As fitDiscrete is only able to use a single chronogram and not a standard error distribution of relative node ages, only the point estimate chronogram is presented.

RESULTS

*Phylogenetic analyses*—Figure 2 shows the maximum likelihood tree generated by PAUP*, with support values generated from the Bayesian (posterior probabilities: PP), maximum likelihood (ML) bootstrap, and maximum parsimony (MP) bootstrap analyses, presented in that order for each node on the
topology. Bayesian PP values are generally very high for nodes relatively deep within the phylogeny, forming a strongly supported backbone that delineates four major clades within subgenus *Limnochloa* (Fig. 2). All four of these clades are present in the ML topology (Fig. 2) and the MP topology (not shown), and all are present in 100% of the trees in the posterior distribution from MrBayes, but MP and ML bootstrap support values vary greatly among them. Likelihood and parsimony bootstrap support are both less than 50% for the branch separating the largest two clades. These clades (A and B) contain the majority of recognized species, while the other two minor clades correspond to the highly polymorphic, widely dispersed species complexes of *E. acutangula* and *E. dulcis*, respectively. The *E. dulcis* clade, which contains several cryptic morphotaxa of the *E. dulcis* complex as well as the morphologically distinct species *E. sphacelata*, is well supported as the sister clade to the rest of subgenus *Limnochloa*, while the *E. acutangula* clade is equally well-supported as the sister clade to the remaining subgenus *Limnochloa* taxa except the *E. dulcis* clade. The outgroups are well-supported in a reciprocally monophyletic relationship with subgenus *Limnochloa*.

**Character reconstructions**—The results of the stochastic mapping and Mk1 ancestral state reconstructions were largely congruent (Fig. 3). The main difference between them is that the Mk1 model reconstructed nodes with greater uncertainty than stochastic mapping when the associated clade contained extant taxa with diverse morphological states. For instance, the node at the base of the clade containing *E. mutata*, *E. sp. nov. guyana*, and *E. spiralis* is ambiguously reconstructed by the Mk1 model, with more than half of all possible states being nearly equally likely. The stochastic mapping reconstruction is less ambiguous, as only the three states present in the tip taxa are considered probable states for the ancestral node. This is also the case for other nodes deep within clade B and the node at the root of subgenus *Limnochloa*, where the Mk1 model assigns similar, relatively high likelihoods to many states, and stochastic mapping assigns higher probabilities to many fewer states.

Transitions among states across all histories generated by stochastic mapping are presented in Fig. 4, which visually summarizes the values from Table 1. An average of 13 transitions among states
occurred per character history. Transitions were highest from round-spongy stems (state 0) to acutely three-angled (also spongy) stems (state 4) and from round-septate stems (state 1) to round-spongy stems. Overall, round-spongy and round-septate stems were much more likely than any other stem type to change to another stem type. The most uncommon stem types (states 2-6) showed some directionality of change, tending to transition to one other type at a much higher rate. The total number of changes to each stem type regardless of the starting type is shown in the bottom-right graph of Figure 4. There were more transitions (c.a. 3) to the acutely three-angled type than to any other type. There were an intermediate number of transitions (c.a. 2) to round-spongy, round-septate with a central vascular bundle, and four-angled stems, and only about one transition per history to round-septate, obscurely three-angled, and five-angled stems.

**Diversification analyses**—The ML value of Pagel’s κ was 0.3503, and the likelihood of the fit of the character data to the topology improved to -47.638 under this scenario compared to the ML value of -49.6327 generated under the null model (i.e., the fit of the data to the topology lacking κ). This represented a significant increase in the fit of the data (likelihood ratio test, \( p = 0.04579 \)). Figure 5 shows the 3-dimensional likelihood surface for this model, magnified to detail the highest peak, where the ML value of κ was found. Figure 6a shows the ultrametric tree generated by r8s, while Fig. 6b shows the effects of modifying this tree by exponentiating all branch lengths to the power κ. The branches of the tree, especially the very short ones near the tips, are much more equivalent in length after the application of κ.

The ML value of δ was estimated to be 1.6538, but this did not provide a significant increase in the likelihood of the fit of the data to the topology.
Table 1. Mean transition frequencies among all states as averaged across the distribution of character histories generated by stochastic mapping. The values of the matrix represent the actual number of each type of change per character history. The max proportion column shows the percentage of changes per history away from any given state to the state it changes to most often. The values from which these proportions are derived are shown in bold. These values suggest an apparent directionality of change for some states.

<table>
<thead>
<tr>
<th>Start state</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
<th>Max proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.83</td>
<td>1.06</td>
<td>0.05</td>
<td><strong>2.04</strong></td>
<td>1.03</td>
<td>0.01</td>
<td>5.02</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td><strong>1.99</strong></td>
<td>0.24</td>
<td>0.20</td>
<td>0.26</td>
<td>0.68</td>
<td>0.24</td>
<td>3.61</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.05</td>
<td>0.16</td>
<td><strong>0.19</strong></td>
<td>0.02</td>
<td>0.01</td>
<td>0.44</td>
<td>43%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.03</td>
<td>0.28</td>
<td><strong>0.45</strong></td>
<td>0.01</td>
<td>0.01</td>
<td>0.79</td>
<td>57%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.05</td>
<td>0.36</td>
<td><strong>0.55</strong></td>
<td>0.02</td>
<td>0.01</td>
<td>1.01</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
<td><strong>0.74</strong></td>
<td>0.95</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td><strong>0.25</strong></td>
<td>0.32</td>
<td>78%</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

*Phylogenetic patterns*—This study represents the first detailed phylogeny for the species of *Eleocharis* subgenus *Limnochloa*. Maximum likelihood and Bayesian analysis of the subgenus uncovered four major clades (Fig. 2). Although Bayesian PP support values for the relationships among these clades are very high, ML and MP bootstrap values are not. A primary reason for this is that the *E. acutangula* clade, though shown in Fig. 2 as sister to all subgenus *Limnochloa* except the *E. dulcis* clade, was placed at the root of clade A in many of the trees generated by the bootstrap replicates. It is possible that *E. acutangula* is more closely related to clade A than it is to other taxa in subgenus *Limnochloa*. Additional taxon sampling may increase confidence for one or the other of these reconstructions.

Clades A and B, as represented in Fig. 2, are generally partitioned according to biogeography. Clade A consists mostly of Old World species, with the notable exception of the South American *E. tiarata* and *E. plicarhachis*, which group together with members of the African *E. variegata* complex. Most of the members of this African/South American clade are relatively small compared to other subgenus *Limnochloa* species and many have red-black margined or reddish-tinged glumes. Based on morphological and biogeographic affinity, we expect that some of the taxa that were not available for including in this study, including the South American species *E. jelskiana* and the African *E. nupeensis*, would also fall within this clade. The rest of clade A is comprised of Australasian taxa, most of which are native to Australia’s Northern Territory, where exceptionally large floodplain ecosystems have provided an ideal opportunity for these plants to diversify (Cowie et al., 2000; Finlayson, 2005). *Eleocharis philippinensis* and *E. ochrostachys* are widespread taxa of the South Pacific that range into southern Asia.

Clade B contains mostly neotropical taxa, but also contains several species that can be found relatively far north into temperate North America (Smith et al., 2002). This clade contains more diversity in stem architecture than the rest of subgenus *Limnochloa* combined (Fig. 3). *Eleocharis yecorensis* has the curious five-pointed star shaped stems found nowhere else in the genus and *E. sp. nov. guyana* has the central vascular bundle morphology known otherwise only from the Australian endemic *E. sp.*
“Coonjimba Billabong entity”. With the notable exception of *E. spiralis*, all the taxa in clade B are endemic to the New World. *Eleocharis spiralis*, however, is endemic to Australia’s Northern Territory. This could represent an interesting case of long-distance dispersal.

Several other species of interest are absent from our phylogenies, including the recently described *E. laeviglumis* R.Trevisan & Boldrini, as well as *E. cellulosa* Torr., *E. kleinii* Barros, *E. liesneri* S.González & Reznicek, and *E. robbinsii*, although it is possible that *E. cellulosa*, *E. kleinii*, *E. laeviglumis*, and *E. robbinsii* may group with *E. elongata* (note that all these taxa are found in the New World, in the most temperate reaches of subgenus *Limnochloa*’s geographic range), we hesitate to speculate further, as the morphology of most of these unincluded taxa is either so ambiguous or so distinct from the currently sampled taxa that numerous plausible scenarios exist regarding their relationships to the rest of the subgenus.

**Character reconstructions**—Understanding patterns of character evolution and their associated processes is of interest to evolutionary biologists for numerous reasons (Ronquist (2004), but in many groups, including sedges, character evolution has been little examined. Most studies concerning the evolution of morphological characters in the Cyperaceae have been restricted to the investigation of the taxonomic and/or phylogenetic utility of such characters, and have made little attempt to address their evolutionary significance. Most of these descriptive studies have focused on *Carex* (Hipp et al., 2006; Starr and Ford, 2001, 1995; Starr et al., 1999; Starr, 1997) and *Cyperus* s. l. (Muasya et al., 2002a, 2002b, 2001). Well-developed studies of character evolution within *Eleocharis* are absent from the literature.

Chromosomal evolution is very diverse in the Cyperaceae, and several recent studies by Hipp (2007) and Hipp et al. (2007, in press) have focused on the evolution of chromosomal structure and its potential connection to processes driving diversification within the genus *Carex*. However, studies examining the evolution of macromorphology in sedges are nonexistent. The unusual diversity of culm structure in subgenus *Limnochloa* suggests that associative patterns may exist in this group between culm characteristics and either (a) ecological adaptations, (b) processes of diversification, or both. These
potential patterns and their physiological basis are of interest to physiologists studying the evolution of ecological adaptations in _Eleocharis_ as well as systematists interested in the processes that drive morphological evolution in sedges.

Our results indicate a complex morphological history for subgenus _Limnochloa_, with numerous instances of morphological convergence and/or character state reversals, and suggest an unusually rapid rate of stem shape evolution within this group. Both of our character evolution simulations suggest that the ancestor of subgenus _Limnochloa_ is likely to have been round and spongy (Fig. 3), but somewhat counter-intuitively, the _E. dulcis_ and _E. actuangulara_ clades, each of which are rooted at one of the two deepest nodes in the phylogeny (Fig. 2), are both comprised entirely of taxa with culms that lack either sponginess (the _E. dulcis_ clade) or roundness (_E. acutangulara_). It is not necessarily surprising though that the deepest nodes in subgenus _Limnochloa_ show transitions away from the round/spongy state, as many other similar transitions state appear to have occurred throughout the evolutionary history of the group (Fig. 3).

Our choice of outgroup taxa certainly prejudiced the reconstruction of the deep nodes in the phylogeny toward a round/spongy state, as the outgroup taxa used bear round/spongy stems. This influence however, is consistent with the phylogenetic history of subgenus _Limnochloa_. The outgroup taxa used here are more closely related to subgenus _Limnochloa_ than any _Eleocharis_ species that have been included in our analyses to-date (data not shown). Furthermore, the round and spongy state is common throughout _Eleocharis_ as a whole as well as within subgenus _Limnochloa_, and even if other closely related taxa were to be used as an outgroup, little to no difference would occur in the reconstruction (results not shown).

Stem structural variation is more conserved in clade A than in clade B. The tip taxa of clade B (Fig. 2) exhibit all the stem architectural types that can be found throughout the entire subgenus (Figs. 1 and 3), suggesting that stem structure within clade B in particular may be evolving at rate significantly faster than the mean rate for the subgenus. In both clades A and B however, where stem architecture does vary it seems to change decisively and unpredictably with respect to phylogeny, and frequent transitions
to morphologically dissimilar states are indicated. Within very recently split lineages, however, some residual similarity can usually be seen between the architectural characteristics of sister taxa. For instance, the *E. quadrangulata* + *E. yecorensis* clade (Fig. 3; nested within clade B) has strongly *n*-angled stems, and *E. spiralis* + *E. mutata* (clade B) have triangular stems, though they differ in degree of acuteness. *Eleocharis equisetoides* and *E. interstincta* (clade B) both have septate stems and *E. philippinensis* shares the character of having strongly *n*-angled stems with the relatively closely related *E. sp.* “Lansen Creek entity” (clade A). Other combinations of closely related taxa however, such as *E. interstincta* + *E. cf. obtusetrigona* (clade B), and *E. sundaica* + *E. sp.* “Coonjimba Billabong entity” (clade A), show stem morphologies with very little in common. The correlation between stem form and phylogenetic relatedness breaks down entirely at nodes more than two branching events deep in the phylogeny, except in the *E. dulcis* clade, where all taxa have round/septate stems, and the *E. variegata* clade (within clade A), where all taxa are round/spongy. However, it would not be surprising if the addition of more taxa were to cause the deterioration of this association within the *E. variegata* clade.

Stochastic mapping and the Mk1 likelihood model are commonly used to reconstruct character histories although differences between the character histories created by these methods occur frequently (Ekman et al., 2008; Renner et al., 2007; Ricklefs, 2007). In our study, the results of the Mk1 and stochastic mapping reconstructions differed in a few clades with tip taxa that had a wide distribution of states (Fig. 3). For the ancestors of morphologically diverse extant clades, stochastic mapping generally created relatively confident reconstructions, in which usually one, but in some case up to three, tip states were considered relatively probable, while the Mk1 model generally created less confident reconstructions, in some cases suggesting an almost equal probability of all potential tip states for ancestors of morphologically diverse clades (e.g. clade B, Fig. 3). Ekman et al. (2008) noticed similar incongruence among these and other recently developed ASR methods in their analysis of lecanoralean lichens. The statistical properties of the models that cause this notable difference in their reconstructions are not well understood (Ekman et al. 2008; Ricklefs, 2007).
Within the context of this study, the high confidence of the stochastic mapping reconstructions has created conveniently intuitive reconstructions, but these reconstructions may not always represent a biologically realistic situation. Stochastic mapping calculates the conditional probabilities of character states at the internal nodes of the tree, and then simulates a character history, starting at the root, using both the stationary frequency of each character state and the states of deeper (i.e., already reconstructed) nodes as factors when determining the state of each node during its simulations (Nielsen, 2002; Huelsenbeck et al., 2003). This can lead to reconstructions with higher confidence in fewer states as compared to those created by methods like the Mk1 model, which calculates the conditional likelihoods of nodes, but does not actually simulate character histories and does not multiply by the stationary frequencies when estimating likelihoods of states at internal nodes. The more ambiguous reconstructions suggested by the Mk1 model may offer a more conservative estimate of the true states of nodes that contain diverse tip taxa, uninformative though it may seem. It is reasonable to assume that if the taxa in subgenus Limnochloa do have a high rate of morphological evolution, ancestors at even relatively shallow nodes may have shown character states that are currently not present in any extant descendents. It is also possible, however, that the Mk1 model is providing an overly conservative estimate, and that the more confident reconstructions of stochastic mapping are biologically sound.

Considering the methodological discrepancies between these recently developed ASR methods and their tendency to produce different inferences of history, it seems wise to follow the advice of Ekman et al. (2008), to employ multiple methods and compare the results, at least until a better understanding of the statistical properties of these methods is developed. Simulation studies examining and comparing the behavior of these recent methods under various biological scenarios (high vs. low rate, assymetrical vs. symmetrical change, etc.) would be useful.

Evolutionary implications of stem shape—One of the major goals of evolutionary biology is the characterization of the relationships between patterns of development, morphology/anatomy, and physiological adaptation. The first step in this process is the characterization of patterns of diversification,
which is what this study had addressed. In *Eleocharis*, these patterns and their ecological significance are not well understood. Despite the fact that *Eleocharis* is a model system for examining photosynthetic adaptations (Ueno, 1988, 2001, 2004; Baksh and Richards, 2006; Murphy et al., 2007; Roalson, 2008), and the main photosynthetic organs in this case are the plants’ culms, there is a general lack of knowledge about the physiological and ecological significance of culm structural characteristics in the genus. It is known that certain stem characteristics, such as Kranz anatomy, are associated with photosynthetic pathway variation (Ueno, 1988, 2001, 2004), but it is not known whether gross morphological characters such as those examined in this study are similarly associated with ecological adaptations.

Studies examining the connections between adaptation and morphology in subgenus *Limnochloa* are rare. Baksh and Richards’ (2006) published a detailed, informative study of the overall morphology, anatomy, and general growth patterns of *E. cellulosa*, but it is primarily descriptive in nature and only superficially addresses questions of adaptive significance. Routledge’s (1987) study of *E. palustris* addresses adaptive significance but does not consider evolutionary implications. Busch et al. (2004) and Chen et al. (2005) examined growth responses of *E. cellulosa* to differing environmental conditions, addressing the adaptive roles of whole-plant growth responses to changes in environment, and documented dissimilar growth patterns between plants subjected to different environmental conditions (submersed vs. emergent; gradients of nutrient concentrations). These studies highlight potential habitat characteristics that could be driving selection for morphological adaptations, but they do not address the potential role of stem architecture in the plants’ responses to these pressures, nor did they characterize any changes in stem structure among plants subjected to differing treatments. A series of experimental studies exploring the dynamics of gas transport in *E. sphacelata* using mathematical simulations that did address the potential adaptive roles of some aspects of stem architecture was undertaken using the Australian species *E. sphacelata* (Sorrell and Boon, 1994; Sorrell et al., 1994, 1997; Sorrell and Tanner, 2000), but these studies were limited to this single species and focused solely on properties of the plants related directly to gas exchange. Thus, although these studies are compelling and may offer insight into the potential adaptive roles of some aspects of *Eleocharis* stem architecture (e.g. stem architecture may affect
gas exchange resistance, which plays a particularly important role in the physiological function of semi- or fully aquatic species), they do not address the variation in stem architecture that is seen throughout subgenus Limnochloa nor Eleocharis as a whole.

Other studies addressing the evolution of form and function in subgenus Limnochloa are absent from the literature, leaving numerous questions unanswered. Of primary interest are the questions (a) whether gross stem shape and internal structure do in fact exert physiological effects on Eleocharis species, (b) the nature (and potential adaptive roles) of these effects, and (c) whether predictable patterns hold among taxa in this group. Even if the effects on fitness of the particular morphological characters we examined were minimal in and of themselves, it would still be useful and interesting to identify whether or not these macromorphological characters were somehow associated with other, perhaps less obvious physiological adaptations. Because of the diversity of its taxa in regard to stem macromorphology, subgenus Limnochloa is a good group for identifying such patterns. A peripheral yet compelling question of interest that remains unaddressed is how any putative patterns observed in Eleocharis might compare to analogous patterns in other monocots. Identifying such broad-scale evolutionary patterns greatly increases our predictive power and furthers our understanding of the evolution of form and function in plants.

The results summarized in Fig. 4 do indicate that there may be some directionality of change for some stem types in subgenus Limnochloa. For example, a taxon with round/spongy stems is twice as likely to transition to acutely three-angled/spongy stems than any other stem type, and this happens relatively often in our simulated histories: twice per tree. However, once a taxon has arrived at the acutely three-angled state, it is much less likely to transition back to a round/spongy state than to another state such as obscurely 3-angled/spongy or round/septate with a central vascular bundle. This directional pattern of evolution is suggestive that some combination of selection and developmental bias is probably acting on traits that are directly or indirectly (e.g., via linkage disequilibrium) associated with stem shape.

The possibility that the effect of developmental bias is stronger than selection in this case cannot be excluded based solely on this directional pattern (Arthur, 2002; Garson et al., 2003; Arthur, 2004;
Brakefield, 2006), but if we also consider the discrepancies among the rates of change in these characters across the lineages of subgenus *Limnochloa*, it seems likely that selection is playing a meaningful role in the evolution of stem shape and architecture, for two reasons: (i) the observed variation in the rate of morphological change is relatively high and seems unlikely to have arisen from chance alone (this is speculation, but could be tested against a null expectation for rate heterogeneity if this could be hypothesized) and (ii) the results of geiger’s best fit test of Pagel’s $\kappa$ parameter to our tree suggest that change in stem structural characteristics is more likely to have occurred at or near speciation events. Even in the shallow subclades of clade B (Fig. 2), stem structural diversity is high (Fig. 3; the number of stem types varies between $n=2$ and $n=3$). The presence of this association in a group like *Eleocharis*, which is well known for its stem elaboration, is highly suggestive that speciation is at least in part facilitated by the adaptability of stem morphology within the group. At the very least, the evidence we present is enough to warrant future studies further testing this hypothesis. The biological significance of this correlation remains unknown. It is not immediately apparent how differences in stem architectural characteristics may generate barriers to gene flow, since selection on stem shape and structure is not understood. Developing a better understanding of the processes that have generated this pattern would significantly increase the resolution of our knowledge regarding the relationship between morphological changes and speciation, particularly in the Cyperaceae and Poales in general.

*Developmental basis of stem structural evolution*—There are numerous studies that address the genetic basis for the evolution of vegetative morphology in angiosperms, and in monocotyledons in particular. Topics of particular interest in the literature have included leaf shape and venation, vascular system structure and development, growth pattern (e.g., sympodial vs. monopodial), photosynthetic adaptations, and issues of homology and reduction/transformation of vegetative parts, such as leaves and branches (Roth-Nebelsick et al., 2001; Shepard and Purugganan, 2002; Tomlinson and Zimmerman, 2003; Givnish et al., 2005; Linder and Rudall, 2005; Piazza et al., 2005; Kidner, 2007). Some studies have addressed the pattern of plant body development as predicted by the behavior of the shoot apical
meristem (David-Schwarz and Sinha, 2007). However, a thorough literature search uncovered no studies directly addressing the developmental or genetic basis of gross stem shape evolution in any land plant lineage, let alone the monocotyledons in particular. The complex directionality of stem structural change found in subgenus *Limnochloa* (Fig. 6) indicates that the developmental genetic basis for stem shape may be similarly complex, and is probably affected by developmental biases inherent in the gene families coding for these patterns. The investigation of these patterns is a potential avenue of future research that would significantly further our understanding of the development of the monocot sporophyte body.

**Concluding remarks**—The phylogenetic history of *Eleocharis* subgenus *Limnochloa* reveals that patterns of stem architectural diversification in this group are dynamic and are not predictably associated with patterns of phylogenetic relatedness within the subgenus. The characters we examined seem to be associated not only with speciation events in this clade, but potentially also with ecological adaptations, although the adaptive significance of these stem structural characteristics remains to be determined. This study represents a first foray into evolutionary character analysis in the sedge family, an area of research that has been largely unaddressed by earlier studies. Future studies of *Eleocharis* stem morphology should focus on the potential adaptive roles of stem traits like (but not limited to) those traits examined in this study, in order to help elucidate the processes driving the evolution of stem structure in *Eleocharis* and the herbaceous monocots in general.


Fig. 1. Stem structural variation in *Eleocharis* subgenus *Limnochloa*. Roman numerals I-V show the variation in gross stem shape across the subgenus. The dashed lines and accompanying shapes show the lateral profile of the culm at that point along its length. In most *Eleocharis* species (including subgenus *Limnochloa*), the culms have the same profile along their entire length. In some species, however, the culms are terete near the base and gradually become trigonous near the apex. Letters A-C show the variation in internal architecture of subgenus *Limnochloa* stems. A and B show septate stems; in A the vascular bundles are restricted to the exterior culm wall, but in stem B a central vascular bundle that extends from the base to the apex of the stem also exists. C shows a “spongy” stem with incomplete transverse septae. In this stem type, vascular bundles are scattered throughout the interior of the culm as well as the exterior culm wall. Each of the seven character states used in the analysis corresponds to a unique combination of one type of gross stem shape (I-V) and one type of internal architecture (A-C). Numbers in parentheses indicate which character state(s) exhibit each type of variation.
Fig. 2. Maximum likelihood tree for *Eleocharis* subgenus *Limnochoa*, found by PAUP*, showing branch support values from the Bayesian analysis (PP), the maximum likelihood bootstrap (ML), and the maximum parsimony bootstrap (MP), in that order. An asterisk (*) represents a value of 100%, while a dash (-) represents a value of less than 50%. Four well-supported major clades are indicated. The clade label “actngl.” designates the “acutangula clade.” Refer to Appendix I for complete species names and authorities.
Fig. 3. Character reconstructions. The stochastic mapping consensus phylogeny was constructed from independent simulations of character histories across all phylogenies in the posterior distribution of trees generated by MrBayes, and the Mk1 model reconstruction was performed on the ML topology determined by PAUP*. Pie graphs on the stochastic mapping reconstruction indicate proportions of character histories in which the ancestor at that node was reconstructed as any given state. Pie graphs on the Mk1 reconstruction indicate fractional likelihoods of the ancestor at that node being in any particular state under the Mk1 model. Node color designations correspond to the stem shapes of the same color in the key below the tree diagram, labeled 0-6. These number labels correspond to the descriptions of these stem types in the text. In the key, hatching represents incomplete transverse and longitudinal septae (i.e.
spongy stems). Stems without hatching have complete tranverse septae and no longitudinal septae. Large and small filled circles represent vascular bundles. Clade labels correspond to labels in Fig. 2. The label “actgl.” designates the “acutangula clade.” Refer to Appendix I for complete species names and authorities.
Fig. 4. Transition probabilities among character states across all character histories generated by stochastic mapping. From top-left, the first seven graphs indicate the frequency of transitions from each character state (0-6) to all other character states. Many character states show a tendency to change more often to some states than others. The graph on the bottom-right shows the summary statistic of the frequency of changes per character history to some given state regardless of starting state.
Fig. 5. Likelihood surface for best-fit value of Pagel’s $\kappa$. The maximum likelihood estimate of the values $\kappa$ and the rate of change parameter $q$ occur within the zone labeled on the graph with a value of 1. Likelihoods of higher values of $\kappa$ decrease slowly up to about a value of 1 (where $\kappa$ has no effect on the shape of the tree) and then rapidly as the value of $\kappa$ increases beyond 1 (where $\kappa$ begins dramatically modifying branch lengths).
Fig. 6. Effect of applying the best-fit value of Pagel’s $\kappa$ to the ultrametric phylogeny. Figure 6A shows the topology generated by r8s, before the application of the maximum likelihood value of $\kappa$. Figure 6B shows the topology with the branch lengths exponentiated to the power $\kappa = 0.3503$. A low value of $\kappa$ homogenizes branch length variation throughout the tree, bringing all branches close to a length of 1, and indicates that most changes in character state are associated with single branching events.
Introduction

Eleocharis R.Br. is a genus of about 250 species with a cosmopolitan distribution. Most species of Eleocharis however, occur in areas that experience a warm, wet climate at least part of the year, and most of the species diversity is concentrated in eastern North America, warm-temperate Asia, and the wet tropics (Svenson, 1929, 1934, 1937, 1939; González-Elizondo & Peterson, 1997; Smith, 2002). Throughout their range, Eleocharis species occur in greatest abundance in riparian or seasonally flooded areas from sea level to sometimes higher than 5,000 m (particularly in the tropical Andes). Numerous species have considerable local ecological or economic significance (Bayliss & Yeomans, 1990; Jordan & al., 1997; Guichon & al., 2003). Eleocharis dulcis (Burm.f.) Trin. ex Henschel, the water chestnut of eastern Asia, is an important agricultural crop grown for its crisp, edible tubers. Several species of Eleocharis show interesting photosynthetic adaptations, such as the ability to switch between C₃ and C₄ carbon fixing pathways to maximize carbon fixing potential in both terrestrial and aquatic environments (Ueno & al., 1986; Bruhl & al., 1987; Ueno & al., 1988; Ueno & al., 1989; Ueno & Samejima 1989; Ueno 1996a, b; Uchino & al., 1998; Murphy & al., 2007; see also Bruhl & Wilson 2007).

The genus Eleocharis is circumscribed by the following morphological synapomorphies: leaf blades are absent or reduced to a mucro and the stems serve as the primary photosynthetic body; inflorescences are solitary, unbranched spikelets at the tips of culms, which may contain 1-many flowers; and the nuts (nearly) always have a persistent style base, which is often referred to as a ‘tubercle.’

Phylogenetic relationships within Eleocharis are poorly understood, with only two phylogenetic studies in the genus (Roalson & Friar, 2000; Yano & al., 2004). Although the genus shows considerable morphological variation, how that variation should be used to define lineages is unclear. Over the last two centuries, various classifications, showing varying degrees of conflict, have been proposed (Torrey, 1836;
Kunth, 1837; Bentham & Hooker, 1883; Clarke, 1900, 1902, 1908; Beauverd, 1921; Svenson, 1929, 1934, 1937, 1939; Zinserling, 1935; Blake, 1939; Koyama, 1961; Egorova, 1976, 1980, 1981; Egorova & Khoi, 1980; Kukkonen, 1990; González-Elizondo & Peterson, 1997). González-Elizondo & Peterson (1997) present the most recent comprehensive revision of the genus. Their classification, which is based on morphological similarity, includes updated circumscriptions for many of the previously recognized supraspecific taxa, and also descriptions of several new classification units. A forthcoming study concerned with the reclassification of *Eleocharis* in light of molecular phylogenetic evidence will address issues of incongruent classifications in much greater detail than is warranted here.

Roalson & Friar (2000) published the first phylogenetic analysis of *Eleocharis* species, using nucleotide sequence data from the nuclear ribosomal DNA internal transcribed spacer region (ITS) to reconstruct phylogenetic history across the genus. They found that many of the taxa delimited by González-Elizondo and Peterson were not monophyletic. A subsequent phylogenetic analysis of Japanese *Eleocharis* species (Yano & al., 2004) corroborated Roalson & Friar’s results. These phylogenetic studies, however, included only sparse sampling within the genus, and left many relationships untested. Current studies are underway to provide an extensive phylogenetic examination of the entire genus (Hinchliff & Roalson, 2009; E.H. Roalson & al., in press).

Of somewhat contentious phylogenetic placement are three monotypic genera, which have variously been considered closely related to *Eleocharis* or included within it. These are (1) the South American submersed aquatic *Egleria fluctuans* L.T.Eiten, (2) *Websteria confervoides* (Poir.) S.Hooper, a submersed aquatic species found throughout the wet tropics and subtropics, and (3) *Chillania pusilla* Roiv., a diminutive terrestrial species with a narrow Andean distribution. *Egleria* and *Websteria* strongly resemble many submersed aquatic *Eleocharis* species, and previous phylogenetic analyses have suggested that at least *W. confervoides* is derived from within *Eleocharis* (Roalson & Friar, 2000). *Chillania pusilla*, known only from a single locality in the Cordillera de Chillán east of Concepción in central Chile, shows all the synapomorphies of *Eleocharis* and falls within the range of variation observed in other species.
Seberg (1985) also observed this and published his assertion that *C. pusilla* should indeed be considered a species of *Eleocharis*. He designated the name *E. uniflora* O.Seberg for this purpose.

Related to the question of where the generic boundaries of *Eleocharis* may lie is the question of how *Eleocharis, Websteria, Chillania, and Egleria* are related to other Cyperaceae taxa. Previous analyses have indicated that the *Eleocharis + Websteria* clade is sister to a clade containing *Bulbostylis*, which by inference we may assume is the Abildgaardieae (Roalson & Friar, 2000). Other family-level phylogenetic studies have suggested *Eleocharis* to be sister to a number of different lineages in the Cyperaceae (among these: *Isolepis, Fuirena, Bolboschoenus, and Schoenoplectus*), but strong support has been elusive (Bruhl 1995; Muasya & al., 1998, 2000a, 2000b, 2001, 2008; Simpson & al., 2007). Goetghebeur (1986) suggested that the Eleocharideae is sister to the Abildgaardieae, and Bruhl (1995; at least in some analyses), based on an analysis of non-molecular data, found a close relationship between *Eleocharis* and Abildgaardieae in at least some analyses. A recent phylogenetic study of the Cyperaceae using *ndhF* (Hirahara & al., 2007) resolved a sister relationship between *Eleocharis* and the Abildgaardieae with strong support. Besnard & al. (2009) also resolved this relationship in their analyses of C₄ sedges.

Another question that has not been made clear by previous studies is where the root of the *Eleocharis + Websteria* clade is attached within the Eleocharideae clade itself. Some analyses have placed the root between *E. subgenus Limnochloa* and the rest of *Eleocharis* (Roalson & Friar, 2000), but this hypothesis has not been resolved with strong support. Determining where this root is located is important for future phylogenetic work within the genus, because of the difficulty of aligning sequence data for loci that are phylogenetically informative within *Eleocharis* to homologous data from outgroups. Once the root has been determined unequivocally, future phylogenetic studies within *Eleocharis* will be able to exclude difficult to align outgroup sequences, which will likely lead to the production of better-resolved trees that will be of more use in downstream comparative analyses (e.g. studies of trait evolution and diversification).

A probable cause for the inconsistency and low support values that characterize previous hypotheses of the rooting of *Eleocharis* is the relatively large phylogenetic distance between the *Eleocharideae* and
its outgroups. Previous phylogenetic analyses using rapidly-evolving markers from nuclear ribosomal DNA (Roalson & Friar, 2000; Yano & al., 2004), which provide good resolution of infrageneric relationships within *Eleocharis*, have nonetheless been plagued by issues of homoplasy in the alignment of sequences of potential outgroups (E.H. Roalson & al., unpublished data). Conversely, very slowly evolving markers such as *rbcL* vary little across genera and are easily aligned. Unfortunately, the variability of these markers is in fact so low that they do not contain enough phylogenetic signal to allow unequivocal inference of relationships among major lineages of Cyperaceae (Muasya & al., 1998, 2000a, 2000b, 2001, 2008; Simpson & al., 2007). We therefore reasoned that to resolve relationships within Cyperaceae at the suprageneric, subfamilial level, it would be necessary to use markers that evolve slower than ITS, but faster than *rbcL*. We chose the chloroplast markers *ndhF* and *psbB-psbH*, as they fit these criteria.

We propose in this study to use chloroplast nucleotide sequence data to address three outstanding questions: (1) what is the phylogenetic placement of *Eleocharis* within the Cyperaceae, (2) what is the phylogenetic position of the genera *Egleria* and *Websteria* in relation to *Eleocharis*, and (3) what are the major clades within *Eleocharis*? Robust phylogenetic hypotheses of these relationships will provide a concrete framework that will be invaluable to future studies concerning the evolution of the Cyperaceae and *Eleocharis* in particular. Our taxonomic names and ranks follow Govaerts & Simpson (2007).

**MATERIALS AND METHODS**

*Plant samples.*—Individuals were selected for sampling from natural (or naturalized) populations collected in Australia, Bolivia, Guyana, Mexico, and the United States, and also from specimens from several herbaria, including K, MO, NY, RSA, US, UT, and WS. Taxa were chosen broadly from clades throughout the Cyperaceae, but sampling was more intense within *Eleocharis* and its likely sister taxon the Abildgaardieae, which contains the genera *Fimbristylis*, *Bulbostylis*, and *Abildgaardia*, among others (Ghamkhar & al., 2007). Although fewer than 10% of extant *Eleocharis* species are represented in this
study, those species that are included were chosen specifically to represent all major clades of *Eleocharis* 
(sensu Roalson & Friar, 2000; Yano & al., 2004; González-Elizondo & Peterson, 1997; E. H. Roalson & 
C. E. Hinchliff, unpublished data), and provide greater than sufficient resolution of the clade to address the 
questions we consider. Efforts were also made to include as many as possible of those genera which have 
recently been hypothesized to be closely related to *Eleocharis*, including *Bolboschoenus, Bulbostyli*, 
*Fimbristyli*, *Fuirena, Isolepis*, and *Schoenoplectus* (Bruhl, 1995; Goetghebeur, 1985; Kukkonen, 1990; 
Muasya & al., 1998).

**DNA sequencing.**—DNA was isolated using a modified 2X CTAB Buffer method (Doyle & Doyle, 
1987). Templates of *ndhF* region (cpDNA) were prepared using a 1:1 ratio of primers “ndhF-A” (5'-TAT 
GGT TAC CTG ATG CCA TGG A-3’) and “ndhF-D1” (5’-CTA TRT AAC CRC GAT TAT ATG ACC 
AA-3’; Olmstead & Sweere, 1994), while templates of the *psbB-psbH* region (cpDNA) were prepared 
using a 1:1 ratio of the primer “psbB-psbH-R” (5’-TTC AAC AGT TTG TGT AGC CA-3’) and “psbB-
psbH-F” (5’-AGA TGT TTT TGC TGG TAT TGA-3’; Xu & al., 2000). Polymerase chain-reaction (PCR) 
amplifications followed standard procedures (Givnish & al., 2005; Kapralov & al., 2006).

Gel electrophoresis was used to verify that each PCR reaction contained only a single product. 
Extraneous primers and free nucleotides in the PCR products were digested using exonuclease and 
Antarctic phosphatase (New England Biolabs, Inc.; procedure available upon request). The resulting 
‘clean’ PCR products were sequenced using the BigDye 3.1 terminator cycle-sequencing reaction 
(Applied Biosystems, Inc.). Cycle sequencing products were cleaned using the Edge Biosystems DTR gel 
purification system and analyzed on an Applied Biosystems 3730 Automated DNA Sequencer (Applied 
Biosystems, Inc.). Sequence chromatograms were proofed, edited, and assembled into contigs using 
Sequencher 4.5-4.8 (Gene Codes Corporation, Inc.), and data matrices containing all sequences were 
aligned using the software MUSCLE v. 3.5 (Edgar, 2005).
**Phylogenetic analyses.**—Phylogenetic hypotheses were constructed under the maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) criteria. Parsimony and likelihood reconstructions were performed with PAUP 4.0b11 (Swofford, 2001). Bayesian analysis was performed with MrBayes 3.1 (Huelsenbeck & Ronquist, 2001). Sequence gaps were treated as missing data.

The heuristic search for the MP tree consisted of $10^4$ random-addition replicate searches, each constrained to allow only $10^7$ topological rearrangements and to save only 1,000 trees with an MP score greater than 4200 (an arbitrary number picked to ensure that only good trees were used for branch swapping). All characters were assumed to have equal weights.

We used the Bayesian Information Criterion (BIC) to select the best-fit model(s) for the ML and BI phylogeny reconstructions. The BIC identifies the model out of the candidate pool with the highest posterior probability of being the true model (Johnson & Omland, 2004). The pool of candidate models we considered comprised all the variations of the general time-reversible substitution model with unequal base frequencies and the parameters $\Gamma$, which describes among-site rate heterogeneity, and $I$, which describes the proportion of invariant sites in the matrix. In its most general sense, this model is GTR + I + $\Gamma$. There are 203 possible substitution matrices under the GTR base model, each of which may be combined with one of four possible combinations of I + $\Gamma$ (1: GTR, 2: GTR + I, 3: GTR + $\Gamma$, 4: GTR + I + $\Gamma$) to yield the 812 candidate models in our pool. To test the fit of the models, a rough ML guide tree was first calculated in PAUP using the combined data matrix and an abbreviated heuristic search. Each of the 812 models in the pool was then optimized on the guide tree once for each of the $ndhF$, $psbB-psbH$, and concatenated ($ndhF + psbB-psbH$) matrices. The resulting -lnL scores for each model were used to calculate the BIC scores.

Different substitution models were selected using the BIC for the a) $ndhF$ data alone, b) the $psbB-psbH$ data alone and c) the concatenated matrix. PAUP does not allow suites of characters to vary independently under ML analysis, so the ML search was performed using the model selected for the concatenated matrix. This model corresponded to a special case of the GTR + I + $\Gamma$ model, with 4
substitution rates, $\alpha = 1.165436$, and $I = 0.254718$. A heuristic search for the ML topology was implemented in PAUP. Neighbor-joining was used to create a starting tree, upon which $10^5$ TBR rearrangements were attempted to find the best tree. Only $10^4$ trees with a $-\ln L$ score greater than 24,000 were kept in memory at any given time, to reduce the time spent branch swapping on suboptimal trees.

Although MrBayes does allow suites of characters to be modeled independently (unlike PAUP using the ML criterion), it does not allow the use of any submodels of the GTR except for 2-rate and 1-rate models. We therefore implemented a separate instance of the 6-rate GTR + I + $\Gamma$ model for the data from each genetic marker, as this represented the closest available model those that were selected for the gene regions individually (both of these were submodels of GTR + I + $\Gamma$). The substitution rates, base frequencies, I, and $\Gamma$ were allowed to vary independently between the $ndhF$ and $psbB-psbH$ instances, but the topology was constrained between them. We executed an MCMC consisting of two concurrent runs, each with 4 chains (default heating parameter value), to $10^7$ generations. The standard deviation of split frequencies at the end of the runs was less than 0.01, indicating that convergence and mixing occurred. A burn-in fraction of 40% was removed from the beginning of the resulting posterior tree distributions.

Maximum parsimony bootstraps were performed in PAUP. Five hundred random-addition heuristic search replicates were performed, each with a rearrangement limit of $10^8$. Maximum likelihood bootstraps were executed in the program RAxML version 7.0.3 (Stamatakis, 2006). The program was instructed to run one thousand bootstrap replicates, using the rapid hill-climbing algorithm. The search terminated at 670 replicates, when the algorithm decided that stationarity had been reached. A GTR + I + $\Gamma$ was used, with all parameters estimated by RAxML, with the distribution of $\Gamma$ discretized into 25 rate categories.

RESULTS

The combined $ndhF$ and $psbB-psbH$ dataset was 2349 base pairs long, and contained 1154 variable sites. Of these, 835 were parsimony-informative.
The results of our phylogenetic analyses are presented in Fig. 7. Support values for most branches in our phylogenetic hypothesis were very high. *Websteria* and *Egleria* are nested within *Eleocharis* with very strong support. This *Eleocharis + Websteria + Egleria* clade (the Eleocharideae) is strongly supported as the sister lineage to the Abildgaardieae (Bayesian posterior probability [expressed as a percent and hereafter abbreviated to ‘PP’]=100, ML bootstrap [MLB]=96, MP bootstrap [MPB]=91; Fig. 7). There is also strong support for the placement of the root of *Eleocharis* on the branch separating subgenus *Limnochloa*, from the rest of *Eleocharis* (PP=100; MLB=100; MPB=100).

Support values along the spine of the tree were also very high, indicating strong support for the monophyly of many major clades within the Cyperaceae. Among the clades with PP, MLB, and MPB scores greater than 99 were: the Abildgaardieae; a clade containing *Cyperus s.l.* and related genera; the Cariceae tribe; and clades corresponding to currently recognized genera including *Rhynchospora*, *Mapania*, *Calyptrocarya*, and *Mesomelaena*.

**DISCUSSION**

*General relationships in the Cyperaceae.*—The *ndhF/psbB-psbH* phylogeny provides a well-resolved and well-supported phylogenetic hypothesis of the major lineages of Cyperaceae, unlike those based on *rbcL* alone or in combination with *trnL-F* or morphology (Muasya et al., 1998, 2000a, 2000b, 2001, 2008; Simpson et al., 2007). Of foremost interest to us, a sister relationship between the Eleocharideae and the Abildgaardieae is strongly supported by the data we present (Fig. 7). Given the similar morphology of the Eleocharideae and Abildgaardieae (Metcalfe, 1971), this relationship is not surprising. A study by Hirahara et al. (2007) using *ndhF* sequence data from Japanese Cyperaceae also identified this relationship. There is equally strong support for the hypotheses that *Isolepis*, *Fuirena*, *Bolboschoenus*, and *Schoenoplectus* are not closely related to *Eleocharis*, despite previous studies which suggested that these lineages might be closely related to *Eleocharis* (Bruhl, 1995; Muasya et al., 1998, 2000a, 2000b, 2001, 2008; Simpson et al., 2007).
Unlike the previous phylogenetic studies of the Cyperaceae based on *rbcL* data, our results indicate a grade of Fuireneae leading to the Cypereae, and that this Fuireneae + Cypereae clade is sister to the Abildgaardieae + Eleocharideae clade (Fig. 7). Fuireneae + Cypereae + Abildgaardieae + Eleocharideae is sister to the Scirpeae + Cariceae clade, and all of these lineages together are sister to the Rhynchosporeae. All these relationships are very strongly supported by our data (PP>99; MLB>95). The earlier diverging lineages (Mapanioideae, Schoeneae, Bisboeckelereae, Sclerieae, etc.) are not well enough sampled here to allow many comments, particularly since some lineages are not represented (e.g., Trilepideae), however, it is clear that *ndhF* and *psbB-psbH* data provide better resolution and support for relationships throughout the tree than previous regions used. Given the clear indications of paraphyly or polyphyly of some tribes and genera (e.g., Schoeneae, Scirpeae, *Scirpus*, *Carex*, *Cyperus*), these classification units need to be reevaluated. Taxonomic revisions of these groups should incorporate strongly supported phylogenetic evidence to further clarify the uncertain relationships within and among these groups.

**Relationships within the Eleocharideae.**—Our results strongly support the phylogenetic rooting of *Eleocharis* on the branch separating *Eleocharis* subgenus *Limnochloa* (including *Egleria fluctuans*) from the rest of *Eleocharis* (Fig. 7). This result has been corroborated by other analyses with much denser sampling (Roalson & al., in press). This knowledge is important for future phylogenetic studies within *Eleocharis*, because it will allow researchers to use rapidly evolving genetic markers that show relatively high variability within *Eleocharis*, without the need to align the resulting highly polymorphic sequences to distantly related outgroups (Roalson & Friar, 2000). The phylogenetic distance between *Eleocharis* and its closest sister group, the Abildgaardieae, is large enough that those markers that provide good resolution of relationships within *Eleocharis* (such as the nrDNA ITS) cannot be unambiguously aligned with species of Abildgaardieae. Conversely, those markers that can be aligned (such as the markers used for this study) do not provide good resolution of relationships within *Eleocharis* (refer to the unresolved clades within *Eleocharis* in Fig. 7). This problem may now be circumvented through the use of subgenus *Limnochloa* as a functional outgroup in phylogenetic studies within *Eleocharis*. 

46
This study provides the first strong support for the relationships between *Eleocharis*, *Egleria*, and *Websteria*. It is clear that both *Egleria* and *Websteria* were derived from different lineages within *Eleocharis* (Fig. 7), although the position of *Websteria* within *Eleocharis* has been somewhat incongruent across studies and is not yet unequivocally indicated (Roalson & Friar, 2000; Roalson et al., 2009).

*Egleria* is most closely related to species in subgenus *Limnochloa*, while *Websteria* appears to be more closely related to species in subgenus *Eleocharis*. Although *Egleria* and *Websteria* share some characteristics associated with adaptation to submerged aquatic growth (e.g., filamentous culms), they differ morphologically in numerous key ways, so this result is not entirely unexpected. In fact, L. T. Eiten noted these differences in her original description of *Egleria* (Eiten, 1964), and hypothesized about the independent origins of these two taxa. She wrote: “*Egleria* can be imagined to have evolved from a branching rhizomatous species of *Eleocharis* which had spaced, erect stem clumps at intervals,” (p. 483) and then elsewhere suggested that “*Websteria* apparently evolved directly from *Eleocharis* also, but from another part of the genus” (p. 483). The results presented here support Eiten’s hypotheses unequivocally.

Characters used to separate *Websteria* and *Egleria* from *Eleocharis* include having filamentous, branching culms (Fig. 9A & 9C), and an obligate submersed-aquatic habit. Other species of *Eleocharis* usually have non-branching culms and grow terrestrially or as emergent aquatics. However, exceptional morphologies have been frequently noticed in *Eleocharis*, and many species are known with proliferous or otherwise branching growth forms that superficially resemble those found in *Websteria* and *Egleria*. Of particular note is *E. glauca* Boeck. (Fig. 9D), a submersed aquatic Amazonian species with dimorphic, branching culms. This species is very similar in overall appearance to *Egleria*, and is frequently misidentified as such. The African submersed aquatics *E. naumanniana* Böeck., *E. monantha* Nelmes, and *E. caillei* Hutch. ex Nelmes (which has been considered a subspecies of *E. naumanniana*) also have branching culms (Fig. 9B) and are frequently misidentified as *Websteria*. Furthermore, *Websteria* itself is frequently confused with other branched species of *Eleocharis* found in the Americas (Bruhl, 2003).

Members of the *Eleocharis minima*/*E. retroflexa* clade (including *E. baldwinii*) and *E. vivipara* often reproduce by proliferation from culm apices (Fig. 10A & 4B) and may appear to show a branching pattern.
that is similar to that of *Egleria* or *Websteria* when they are growing submerged, although it is important to recognize that the culms of the *E. minima* and *E. vivipara* clades lack the strong dimorphism that is shared by the others mentioned above. It should also be noted that the abilities of all these *Eleocharis* species to either a) create the dimorphic, filamentous culms that appear to confer some advantage in aquatic habitats or b) form proliferous ramets from culm apices, may have been independently derived in numerous unrelated lineages across the *Eleocharis* phylogeny (Roalson & al., in press).

Goldberg and Igić (2008) showed strong evidence that hypotheses of directional character evolution cannot be satisfactorily disentangled from the alternative scenario in which differential rates of diversification are associated with character states, without explicit statistical testing using highly parameterized models such as BiSSE (Maddison & al., 2007; FitzJohn & al., 2009). This argument can be easily extended to cover any kind of character evolution (not just unidirectional change) where diversification rates are assymetrical among states (Goldberg & Igić, 2008; Igić, B., unpublished data). Since it is entirely plausible that the vegetative proliferation character may be associated with a differential rate of diversification in lineages that show it, future work addressing of the question of how and where the proliferation character has evolved in *Eleocharis* should necessarily involve analyses such as BiSSE that take this possibility into account.

*Websteria* has also been separated from *Eleocharis* based on its single-flowered spike and lack of a well-developed tubercle (González-Elizondo & Peterson, 1997). However, several species of *Eleocharis* have single-flowered spikes (*E. naumanniana*, *E. monantha*, *E. caillei*, *E. tucumanensis* Barros, and *E. capillacea* Kunth, among others), and there are several separate instances of tubercles confluent with the nut (e.g. “*Pauciflorae*” sensu Svenson, 1929; González-Elizondo & al., 1997; Roalson & Friar, 2000).

While we were not able to include a sample of *Chillania* in the phylogenetic analyses presented here, Seberg (1985) made a strong case for the inclusion of *Chillania pusilla* within *Eleocharis*, and we concur with his assessment. Morphologically, *Chillania* is much more similar to *Eleocharis* than either *Egleria* or *Websteria* (Fig. 10C) and easily falls within the range of variation found among *Eleocharis* species. This is particularly apparent in the assessment of one of the main characters supposedly differentiating
Chillania from Eleocharis: the presence of four non-peripheral vascular bundles (Roivainen, 1933). When this character is compared with the vasculature of other diminutive Andean Eleocharis species, there is little difference. This is easily seen by comparing the drawing of the stem anatomy Roivainen provides with the photograph of transverse stem anatomy of Eleocharis tucumanensis (Fig. 10D, reproduced from Guaglianone & al., 1998). Eleocharis tucumanensis has 3 or 4 vascular bundles occurring within the stem in a pattern that is apparently identical to that observed in Chillania (Fig. 10C & 10D).

Implications for classification.—The results presented here strongly support the conclusion that Eleocharis is paraphyletic in relation to Websteria and Egleria. This assessment is also supported by Eiten’s comments in her original description of Egleria (Eiten, 1964). We are strong proponents of monophyletic genera and, morphologically, these segregate genera fall well within the range of variation found within Eleocharis. Given the nature of the character states used to separate Chillania and Eleocharis, it is very likely that C. pusilla is also nested within Eleocharis. We therefore suggest the formal recognition of these three species within Eleocharis. Our proposed nomenclature, including all known nomenclatural synonyms, follows. Several nomenclatural synonyms of E. confervoides do not have a designated type or the type material is unclear. These names will require further study and possible future lectotypification.

Dulichium confervoides (Poir.) Alston in H.Trimen, Handb. Fl. Ceylon 6(Suppl.): 310. 1931
Thouars (P [Th]).


= *Scirpus natans* Griseb., Cat. Pl. Cub. 238. 1866, nom. illeg.


= *Scirpus ruppioides* Thw. ex C.H.Wright., Fl. Cub. 176. 1871 – Type: ?.


In recognizing *Eleocharis confervoides*, several nomenclatural issues should be addressed. Five authors have been attributed (in one way or another) with the combination *Eleocharis confervoides*: Kunth (1837), Steudel (1855), Miquel (1859), Koyama (1985), and Tucker (1987). However, none of these authors directly comment upon the combinations made by other authors (other than those referencing Kunth, see below), nor provide a rationale for the combination. If we start with Steudel, he provides the combination and cites the basionym, but attributes the name to Kunth (“Kunth. (Cyp. 113)”), and interjects a question mark between the generic and specific epithets, invalidating this combination (ICBN Article 34.1(a)). Kunth, however, did not make this combination, but rather notes at the end of his treatment of *Scirpus confervoides*, “Eleocharidis species?” Further, this treatment is on page 173, not 113, as cited by Steudel. Similar to Steudel, or following his lead, Miquel attributes *Eleocharis confervoides* to Kunth (“Kunth *l.c.p. 173*”), and also includes a question mark between epithets, invalidating this as a valid combination in *Eleocharis*. Much later, Koyama (1985) makes the valid combination in his treatment of *Eleocharis* in the Flora of Ceylon, likely unnoticed by Tucker (1987) who made the same combination two years later. Also of note, Hooper (1973) suggests that *Chillania*, *Egleria*, and *Websteria* should be treated in *Eleocharis*, despite the fact that she had made the valid combination as *Websteria confervoides* only a year before (Hooper, 1972), but does not cite any publication nor authorship for these nomenclatural changes. González-Elizondo & Peterson (1997) properly cite Koyama as the author of the *Eleocharis* combination, but do not discuss the other names.

**Eleocharis fluctuans** (L.T.Eiten) E.H.Roalson & C.E.Hinchliff, **comb. nov.** *Egleria fluctuans*


Note. While the types have not been seen due to a lack of availability, the paratype, *A. Ducke 12028*
(K), has been studied.


en fylogenetse van de Cyperaceae-genera. Thesis, Rijksuniversiteit Gent [In Flemish].

2741.

González-Elizondo, M.S. & Peterson, P.M. 1997. A classification of and key to the supraspecific taxa 

González-Elizondo, M.S., Peterson, P.M., & de la Cerda-Granzow, I. 1997. A phenetic and cladistic 

Govaerts, R., & D.A. Simpson. 2007. *World Checklist of Cyperaceae*. Royal Botanic Gardens, Kew, 
London.


*Limnochloa* (Cyperaceae): evidence of rapid dynamic evolution in a group of pantropical sedges. 

based on DNA sequences from chloroplast *ndhF* and 5.8S nuclear ribosomal DNA. *Acta Phytotax.* 
*Geobot.* 58: 57–68.

Hooper, S.S. 1972. New taxa, names and combinations in Cyperaceae for the ‘Flora of West Tropical 


55


eexpression is not directly dependent on Kranz anatomy in an amphibious sedge Eleocharis


Ueno, O. 1996b. Immunocytochemical localization of enzymes involved in the C3 and C4 pathways in the

Ueno, O., & Samejima, M. 1989. Structural features of NAD-malic enzyme type C4 Eleocharis: an
additional report of C4 acid decarboxylation types of the Cyperaceae. Bot. Mag., Tokyo 102: 393–
402.

Ueno, O., Takeda, T., & Murata, T. 1986. C4 acid decarboxylating enzyme activities of C4 species

amphibious plant, Eleocharis vivipara: expression of C4 and C3 modes in contrasting

Ueno, O., Samejima, M., & Koyama, T. 1989. Distribution and evolution of C4 syndrome in Eleocharis,
a sedge group inhabiting wet and aquatic environments, based on culm anatomy and carbon

regions of chloroplast DNA of soybean and related wild species and its implications for the

Eleocharis (Cyperaceae) based on ITS sequence data, and chromosomal evolution. J. Pl. Res.
117: 409–419.

URSS 3. Leningrad.
Figure 7: A subtree of the maximum likelihood (ML) phylogram (excluding the Abildgaardieae, Eleocharideae, Cyperaceae, and Fuireneae) inferred from \textit{ndhF} and \textit{psbB-psbH} data. Fig. 8 shows the phylogeny of the excluded taxa, which form a monophyletic group rooted on the branch labeled ‘Fig. 8.’ Support values are indicated for each node in the tree, in the following format: PP / MLB / MPB where PP = posterior probability (scaled to a percentage), MLB = ML bootstrap percentage, and MPB =
maximum parsimony bootstrap percentage. Asterisks (*) indicate a value greater than or equal to 98 for that node, while dashes (-) indicate a value less than 50. Nodes supported by PP, MLB, and MPB values greater than 98 are labeled with a single asterisk. Nodes not bearing any support values greater than 50 are not labeled.
Figure 8: A subtree of the maximum likelihood (ML) phylogram (excluding the taxa contained in Fig. 7) inferred from \textit{ndhF} and \textit{psbB-psbH} data. The subtree shown here is rooted on the branch in Fig. 7 labeled ‘Fig. 8.’ Node labels follow the format used for Fig. 7. All the taxa within the \textit{Rhynchospora} clade
represent species of *Rhynchospora*, although some of their names have been abbreviated as *R. [specific epithet]* to save space within the image.
Figure 9: Photographs showing important morphological characteristics of A: Websteria confervoides (Poir.) S. Hooper, B: Eleocharis monantha Nelmes, C: Egleria fluctuans L. T. Eiten, and D: Eleocharis glauca Böeck. Photographs represent the following herbarium voucher specimens. A: Rhodesia, 1961. E. A. Robinson 4245, K; B: Zaire, Shaba Province, 1975. S. S. Hooper & C. C. Townsend 561, K; C: Brasil,
Pará, 1912. A. Ducke 12028, MICH; D: Venezuela, Guárico, 19981. F. Delascio, R. Montes, & G. Davidse 11416, NY.
Figure 10: Photographs and diagrams showing important morphological characteristics and overall growth habit of A: *Eleocharis vivipara* Link, B: *Eleocharis retroflexa* (Poir.) Urb., C: idealized line drawing of the transverse stem section of *Chillania pusilla* Roivainen, and D: idealized drawing of the transverse stem section of *Eleocharis tucumanensis* Barros. Figure 10C was created from an image (scale not provided).
previously published by the Finnish Zoological-Botanical Society; Figure 10D was created from an image originally published by the Instituto Botanico Darwinion of San Isidro, Argentina. Photographs A and B represent the following herbarium specimens. A: United States, 1880. J. D. Smith s.n., US; B: Central African Republic, Manovo-Gounda-St. Floris National Park, 1983. J. M. Fay 5807, MO.
INTRODUCTION

The sedge family, Cyperaceae, is one of the world’s ten most speciose families of flowering plants (Stevens 2008), containing over 5,400 species (Govaerts et al. 2007). Sedges are most diverse in the tropics, but many lineages have radiated in temperate regions, making the family an ideal model system for exploring patterns of global geographical diversity in plants. Sedges are found in nearly all plant-supporting habitats on all continents except Antarctica, spanning arctic tundra, alpine meadows, temperate and tropical forests, savannas, prairies, marshes, swamps, and deserts (Bruhl 1995; Dahlgren et al. 1985), and are also very phenotypically diverse, including tiny ephemerals, fire-resistant tussocks, scandent vines over 10 m long, and submerged aquatic herbs. These characteristics also provide opportunities for exploring the effects of various ecological traits and processes on patterns of diversity, making sedges a compelling system for the exploration of macroevolutionary ecological dynamics in plants.

In this paper, we use supermatrix methods to reconstruct sedge phylogeny in order to take advantage of as much data as possible and thereby maximize phylogenetic power. To this end, we constructed a large combined matrix of nucleotide data for sedges from the GenBank database, using a novel data pipeline that relies on the software phlawd (Smith et al. 2009; Smith and Donoghue 2008). Phlawd gathers data from GenBank, and automates recursive profile alignments in an attempt to maximize site homology in the resulting matrices, even at deep evolutionary scales. The pipeline presented in this paper further improves the utility of these alignments through a series of complementary refinement steps that filter taxa and genes to 1) filter out tip nodes with questionable data integrity, in order to improve tree resolution, and 2) improve the partial decisiveness (Sanderson et al. 2010; Steel and
of the matrices. Partial decisiveness is a useful metric for describing the limits of a multigene dataset to inform phylogenetic inference (Sanderson et al. 2010; Steel and Sanderson 2009); it describes the number of potential branches in the universe of all possible trees that may be resolved given the pattern of missing data in a given alignment. Most matrices with large amounts of missing data have a partial decisiveness of less than 1, indicating that some branches simply cannot be reconstructed regardless of the phylogenetic quality of the data that are present. Maximizing decisiveness can therefore generally be expected to improve tree resolution in the majority of real-world scenarios.

Supermatrix methods offer a variety of advantages, including, perhaps most significantly, the ability to reconstruct phylogeny at very broad scales with minimal investment in sequencing (Davis et al. 2010; Wolsan and Sato 2010; van der Linde et al. 2010; Ren et al. 2009; de Queiroz and Gatesy 2007; McMahon and Sanderson 2006). With highly resolved trees, it may also be possible to explore broad-scale macroevolutionary patterns with greater confidence than smaller, less resolved trees can provide (Holton and Pisani 2010; Smith et al. 2009; Smith and Donoghue 2008). However, these methods often present their own challenges, including issues of computational power and time efficiency, unanswered questions regarding the effects of the highly incomplete alignments that are characteristic of these approaches (Wiens and Morrill 2011; Lemmon et al. 2009), as well as concerns about data integrity and validation. We explore these questions in the context of our study and also consider the role of some recent advances in phylogenetic theory such as data decisiveness (Sanderson et al. 2010; Steel and Sanderson 2009), and we present and discuss some potential solutions to improve the efficacy and reliability of these powerful new methods.

**METHODS**

*Data collection and validation*—Data were gathered from GenBank release 185 using the software tool phlawd (Smith et al. 2009), which searches the NBCI database for all nucleotide sequences within a given NCBI-recognized taxon, that match a given text query. For instance, it is possible to
generate an alignment of all NCBI sequences matching the query string “internal transcribed spacer” within the Cyperaceae. Because some non-orthologous sequences may be returned for any given query, phlawd requires a set of pre-supplied guide sequences for each query. Any taxa that diverge from these guide sequences by more than some arbitrary amount of coverage and identity are excluded. We used cutoff proportions of 0.3 for coverage and 0.2 for identity.

We gathered data on 23 genes for all available species within the Cyperaceae (Fig. 1), generating alignment files representing over 1,500 species from almost all the family’s genera. Guide sequences for each gene were chosen arbitrarily from the set of available GenBank data, in a manner that sought to maximize their phylogenetic coverage. The resulting alignment files were concatenated on the basis of species name. This approach combines available data from multiple exemplar specimens of each species in order to make phylogeny estimation at this scale possible, and it has proven effective for estimating phylogeny of large numbers of species when parallel data from multiple markers is not available from a single specimen (Smith et al. 2009; Smith and Donoghue 2008; Jones et al. 2002).

Because data integrity on public databases such as GenBank is uncertain, it is possible to include poor or incorrectly identified sequences when data-mining these resources. When multiple sequences from different vouchers are concatenated on the basis of species identity, the potential exists for sequences from misidentified species to be concatenated with correctly identified ones, leading to the creation of so-called “chimeric” taxa. The phylogenetic tip nodes that are associated with these mistaken taxa are expected to contain conflicting phylogenetic signal which, at least in the case of bootstrapping, can erode support values for otherwise well-supported clades, because the chimeric tip is either expected to be placed near each of its different constituent species in different replicate searches, or to contain strong enough conflicting signal that it simply causes parts of the tree to collapse. Automated methods to identify these taxa a posteriori score each taxon for a statistic representing its phylogenetic stability or “leaf stability,” that is, how much its position changes relative to other taxa in replicate tree searches. Those taxa with the lowest stability are assumed to contain either conflicting phylogenetic signal (especially in the case of chimeric taxa) or such low levels of phylogenetic signal that many placements
are equivocal. These unstable taxa are often referred to as “rogue” taxa, and they are frequently removed from further tree searches in order to maximize the credibility of the resulting hypotheses (Thomson and Shaffer 2010; Thorley and Wilkinson 1999).

Multiple methods of measuring leaf stability have been proposed (Smith and Dunn 2008; Thorley and Page 2000; Thorley and Wilkinson 1999). For this paper, we built upon previous work by Maddison and Maddison (2010) in Mesquite version 2.7. They present a statistic, $I$, which is a summary metric of taxon movement among a set of trees, based on differences in patristic distance between all taxon pairs across all pairs of trees in the set. Each difference in distance for a pair of taxa $i$ and $j$ between each pair of trees $x$ and $y$ is scaled to the sum of the patristic distance between $i$ and $j$ across both trees $x$ and $y$, and these are summed across all trees. One drawback (or benefit) of the Mesquite implementation of this method is that it requires a character matrix to be loaded because the instability scores for each taxon are plotted against the percentage of missing data for that taxon in the alignment. Another is that because the magnitude of the scores depends on the number of trees and the number of taxa, direct comparison of $I$ scores generated from different distributions of trees is not straightforward. It is also unfeasible to perform these analyses in Mesquite using very large phylogenies because of memory restrictions. To circumvent these challenges of calculation and interpretation, we wrote a Python program that can take advantage of parallel multiprocessing architectures and large amounts of memory in order to rapidly calculate instability scores even for large sets of very large trees, and which generates similar but differently scaled $I^S$ scores, which are scaled instead to the random expectation for taxon movement in the provided set of trees. Thus, for all combinations $(x, y)$ of trees $x$ and $y$ in some set of trees $X$ (e.g. a set of bootstrap replicates or a Bayesian posterior distribution), and all taxa $j \neq i$ in these trees, the summary statistic $I^S_i$ for a given taxon $i$ is calculated as

$$I^S_i = \frac{\sum_x \sum_j [D_{ij} - D_{ij}] \sum_{x \neq j} D_{ke} \cdot n}{D_{ke} \cdot n}$$
where $D_{ijx}$ is the unweighted patristic distance between taxa $i$ and $j$ in tree $x$, $D_{ijy}$ is this distance in tree $y$, $D_{Re}$ is the random expectation for this distance, which is calculated by taking the average per-tree value of the expression in the numerator on a subsample of trees from $X$ with their tips randomized, and $n$ is the number of trees in $X$. This statistic has the advantage of being directly interpretable as a multiple of the random expectation for taxon movement, and as such it may also be compared directly even between different distributions of trees with different sets of taxa.

*Data filtering and heuristic searches*—We used two strategies to subsample the data gathered by phlawd in order to improve the resolution of resulting trees. The first strategy consisted of removing the top 10% of taxa with the highest $I^S$ scores in order to improve branch stability, and the second of filtering taxa lacking sequences for both of the two most broadly sampled and phylogenetically informative loci (at least for deep nodes) in the dataset, in this case *rbcL* and *ndhF*. This second approach, which we refer to as phylogenetic “scaffolding,” improves the partial decisiveness (Sanderson et al. 2010; Steel and Sanderson 2009) of the matrix by maximizing taxon coverage for the elected markers, and relies on the assumption that those markers contain sufficient signal and sampling to reconstruct major relationships (i.e. deep nodes) in the tree.

To assess the relative utility of both of these approaches, we conducted four parallel rounds of inference using datasets filtered by either 1) neither method (i.e. raw data), 2) filtering rogue taxa only, 3) filtering taxa lacking sequences for elected loci only, or 4) both methods (Table 1). We performed maximum likelihood bootstraps to estimate phylogeny from these alignments, using RAxML version 7.2.6 (Stamatakis 2006). For each alignment, a 300-replicate rapid bootstrap heuristic search (RAxML’s “-f a” algorithm) was used to search treespace. Finally, we calculated descriptive statistics about the alignments and the resulting bootstrap trees to compare the efficacy of our filtering methods for improving the resolution of resulting trees.
RESULTS

A heatmap of sampling density for select markers across all currently recognized Cyperaceae genera is presented in Figure 1. Only a small handful of loci (those at the top of the matrix) show reasonably broad sampling across the family. The broadest sampled are \textit{ndhF} and \textit{rbcL}. The second to last row, ABS\textsubscript{importance}, contains values of a statistic we developed to quantify taxon coverage for each genus, for the markers used in this study (see Appendix I for formulas). Because all genera have been poorly sampled for most of these markers, all their importance scores are quite similar. We therefore calculated the percentile rank of each importance score, which are presented in the last row, ABS\textsubscript{import\_rank}. High scores in this row indicate those genera with the lowest levels of sampling. This heatmap was generated using the gplots package in R (Warnes et al. 2011).

Statistics describing the alignments and trees resulting from each of the four possible combinations of filtering techniques are presented in Table 1. Each subsequent row in the table represents an increase in the stringency of the filtering method. Each of these increases was associated with an overall improvement to the resolution of the resulting consensus trees (Fig. 2), as evidenced by the increase in the proportion of resolved nodes (Table 1). Although the frequency of nodes with greater than 0.6 bootstrap proportion (BP) varied little as filter strictness increased, the proportion of nodes with 0.7 or greater BP showed a steady increase. The most dramatic improvement was seen when taxa not represented by either an \textit{rbcL} or \textit{ndhF} sequence were excluded from the alignment—the proportion of nodes resolved with greater than 0.95 BP increased by 8%, with a concomitant and presumably related 10% increase in the decisiveness of the matrix. The best-resolved and best-supported tree (Figures 3-6) was generated from the most strictly filtered alignment (Table 1, row 4), which contained only the 90% most stable tips from the alignment consisting only of those taxa represented by a sequence for at least one of the loci \textit{ndhF} and \textit{rbcL}.

The majority-rule consensus topologies from the ML bootstrap searches performed on the alignments presented in Table 1 are presented in Figure 2. The trees are presented without taxon names or
support values because space constraints preclude legible presentation of these data in printed form. An overall increasing trend in the proportion of resolved nodes that accompanied more stringent filtering in this study can clearly be seen in this figure.

The best-resolved consensus tree, corresponding to the most heavily filtered alignment (Table 1, row 4), is presented in Figures 3–6. Support values on this tree are generally high, especially for deep nodes.
Table 2. Descriptive statistics for alignments and trees resulting from each of four combined approaches to data filtering.

<table>
<thead>
<tr>
<th>Method:</th>
<th>Tips</th>
<th>Sites</th>
<th>Missing</th>
<th>Decisiveness</th>
<th>Resolved</th>
<th>&gt; 0.95</th>
<th>&gt; 0.7</th>
<th>&gt; 0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All loci, all tips</td>
<td>1,526</td>
<td>20,017</td>
<td>0.89</td>
<td>0.74</td>
<td>0.57</td>
<td>0.31</td>
<td>0.72</td>
<td>0.86</td>
</tr>
<tr>
<td>2. All loci, 10% unstable tips</td>
<td>1,366</td>
<td>19,425</td>
<td>0.88</td>
<td>0.76</td>
<td>0.64</td>
<td>0.33</td>
<td>0.74</td>
<td>0.87</td>
</tr>
<tr>
<td>filtered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Select loci, all tips</td>
<td>484</td>
<td>16,025</td>
<td>0.79</td>
<td>0.86</td>
<td>0.71</td>
<td>0.41</td>
<td>0.75</td>
<td>0.87</td>
</tr>
<tr>
<td>4. Select loci, 10% unstable</td>
<td>435</td>
<td>16,016</td>
<td>0.79</td>
<td>0.86</td>
<td>0.76</td>
<td>0.42</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>tips filtered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rows correspond to filtering approaches: 1. No filtering, all available data for the markers chosen at the data-mining step are used. 2. Filtering of unstable tips only—all sequence data from the 90% most stable tips are used. 3. Filtering of less informative loci only—taxa not represented by at least an *rbcL* or *ndhF* sequence are removed, all other available markers for the remaining taxa retained. 4. Filtering of loci and taxa—all available loci for the 90% most stable taxa represented by at least *ndhF* or *rbcL* are retained. All decimal values are proportions. The last three columns represent the proportion of resolved nodes bearing a bootstrap support value greater than or equal to the respective cutoff value.
DISCUSSION

*Implications for large combined analyses*—Results from our large tree analyses are promising. The best resolved trees from this study have higher support values overall and a greater proportion of well-resolved nodes than the less inclusive family-level trees from the studies which generated the data that we used (Hinchliff et al. 2010; Muasya et al. 2009; Katsuyama et al. 2007), suggesting that data mining approaches such as the those employed in this study may offer substantial advantages over the generation of novel sequence data, as long as these data are already freely available in public databases. However, it is clear from our results that taking advantage of data mining requires a thoughtful approach to ensure data integrity and to optimize phylogenetic resolution. Not only is it necessary to validate taxon placement by eye in order to identify potentially misidentified sequences, but it may also be necessary to filter the available data to achieve the best possible results.

The “scaffolding” approach we used, wherein two well-sampled loci were used to form an alignment “backbone” onto which other available sequence data were concatenated, has the advantage of ensuring the ability to reconstruct deep nodes (using the backbone loci) while allowing the inclusion of numerous, faster evolving loci that can inform phylogenetic relationships closer to the tips of the tree, even when sampling for these loci does not overlap among clades. The nature of modern phylogenetic inquiry has generated aggregate datasets to which two important conditions typically apply: 1) because of the sampling efforts of genus- or lower-level studies, many closely related species are frequently represented by sequences at the same rapidly evolving loci; and 2) because of family- or higher-level studies, a few, broadly sampled and distantly related exemplar taxa are frequently represented by more slowly evolving markers. These aggregate datasets lend themselves well to the scaffolding approach.

Identifying loci to serve as the scaffold backbone requires consideration not only of how phylogenetically informative these loci are, but also how broadly sampled they are across the phylogeny. For most species rich taxa that have been the focus of past phylogenetic inquiry, ample data are available from one or two loci meeting the dual criteria of 1) phylogenetic informativeness (López-Giraldez and Townsend 2011;
Townsend and López-Giráldez 2010) for deep nodes in the tree and 2) broad sampling across the tree. We chose to filter our alignments based on *ndhF* and *rbcL* because these two markers provide adequate signal to resolve many deep nodes within the tree, and for this reason they have been broadly sampled across the Cyperaceae in previous family level studies (Hinchliff et al. 2010; Muasya et al. 2009; Katsuyama et al. 2007).

Filtering out the 10% least stable taxa from the most-inclusive alignment (Table 1, row 2) resulted in an 7% increase in the proportion of resolved nodes, whereas filtering the 10% least stable taxa from the locus restricted alignment resulted in a 5% increase. These differences may seem small, but visual inspection of the resulting consensus trees (Fig. 2, compare A vs. B and C vs. D) confirms that they represent a meaningful increase in resolution. The newly resolved nodes represented by the percent increases mentioned above are concentrated in areas of the tree that are mostly to entirely unresolved in the consensus trees generated using the unfiltered alignments. So although the overall percent contribution of these newly resolved nodes is small, they represent a significant improvement for the phylogenetic information content of the resulting trees. Some areas of the tree actually experienced an apparent decrease in resolution as a result of stringent filtering of rogue taxa between the trees in Fig. 2C and Fig. 2D (see dotted outlines in this figure). However, the nodes that were lost were characterized by relatively low support values, and with their removal, support values for surrounding nodes increased. In some cases the nodes present in Fig. 2C but absent from Fig. 2D were known to mismatch with classifications and other strongly supported studies (for instance, this is the case for a node in Fig. 2C that split the otherwise universally recognized Mapanioideae clade into a basal grade), suggesting that although parts of the topology in Fig. 2 appear better resolved before rogue taxon filtering, this may in fact represent the effects of data artifacts rather than biological reality.

Several approaches to identifying leaf instability exist (Maddison and Maddison 2010; Smith and Dunn 2008; Thorley and Page 2000; Thorley and Wilkinson 1999), and it is unclear which, if any, are ideal. They all serve a similar purpose and should have essentially similar results when used in a filtering step, but their interpretations may be somewhat different. The statistic developed by Thorley and
Wilkinson (1999), implemented in RadCon (Thorley and Page 2000) and subsequently in phyutility (Smith and Dunn 2008), is based on branch support values and as such it does not directly measure the distances moved by taxa among trees. The \( I \) statistic employed by Maddison and Maddison (2010) in Mesquite is based on differences in patristic distance and therefore does directly measure taxon movement, but because it is not standardized, \( I \) scores from different sets of trees cannot be directly compared, and it is difficult to interpret the actual meaning of the summarized taxon scores. The instability statistic \( I^S \) that we employed here has the advantage of creating scores that are directly comparable among sets of trees of differing sizes and with different numbers of tips, and because these scores are scaled to the random expectation for taxon movement, they are straightforward to interpret. A score of \( I^j = 7 \), for instance, means that taxon \( j \) moves seven times more relative to all the other taxa in a given set of trees than the random expectation given that set of trees. A taxon with a score of 14 would therefore move twice this much. The python script we developed to implement the \( I^S \) calculations may additionally be applied to larger sets of much larger trees than other methods allow, as long as adequate computational resources are available.

Partial decisiveness (Sanderson et al. 2010; Steel and Sanderson 2009) is a useful metric for supermatrix studies because it provides a general method of assessing the effect of missing data on tree reconstruction methods, and maximizing partial decisiveness will doubtless prove a useful approach for improving phylogenetic resolution in many cases. However, achieving complete decisiveness is not necessarily a requirement for resolving the so-called “true” tree. Having partial decisiveness less than 1 does not necessarily indicate low phylogenetic utility, because under the common assumption that there is only one, or even several true trees, only a tiny fraction of the universe of all possible branches are represented in those trees. So, as long as these “true” branches can be resolved, very good trees may be found using matrices with partial decisiveness considerably less than 1. The correlation of the partial decisiveness metric with tree resolution in empirical datasets has yet to be fully explored, although we
have demonstrated here that filtering taxa to improve taxon coverage for one or two important markers can improve both partial decisiveness as well as phylogenetic resolution.

*Sequence data sampling strategies*—Although filtering taxa out of supermatrices of course reduces the inclusiveness of the resulting phylogenies, it is a tradeoff that may offer significant advantages in terms of increasing resolution. It is worth pointing out that even heavily filtered supermatrix datasets are often considerably more inclusive than standard datasets and are available for a fraction of the cost and time investment. In the case of this study, our best-resolved tree, with 435 tips, contained only 29% of the Cyperaceae species and subspecific exemplar taxa available on GenBank, but this still represents and addition of 173 tips (a 66% increase) when compared to the next largest published family-level phylogenetic hypothesis of Cyperaceae (Muasya et al. 2009). In addition, the consensus trees presented in this paper indicate stronger support for deep relationships in the Cyperaceae than any previously published tree (Hinchliff et al. 2010; Muasya et al. 2009; Katsuyama et al. 2007), likely due to the greater number and variety of loci incorporated into this study as compared to the others.

The sampling patterns visible in Figure 1 reflect a general phylogenetic sampling strategy that is present in many other groups of organisms as well, and which may be broken down into two complementary approaches. Most studies either A) focus narrowly on a single (often large, species rich) clade, which is typically exhaustively sampled for one or more relatively rapidly evolving markers that may or may not be shared by other studies, or alternatively B) take a broader, less comprehensive sample across a deeper taxonomic scale for one or more relatively slowly evolving markers that are not shared with the finer-scale studies. In the case of Cyperaceae, *ndhF* and *rbcL* have been sampled broadly by family-level studies, while ITS, ETS, and a variety of chloroplast markers have been sampled in a more targeted fashion by infrageneric studies. Because of the low levels of overlap among the markers chosen for different studies, Figure 1 contains a very high proportion of blue cells representing nonexistent datasets for most locus-taxon combinations. However, this apparent lack of sampling is not problematic for phylogeny reconstruction because the nature of the markers chosen is such that they provide adequate
phylogenetic information at the correct time scales and locations in the tree, due specifically to their careful selection by the authors of the studies from whence they originally came.

While it is generally advisable to sequence additional samples in order to improve taxon coverage, it is unnecessarily redundant to expect exhaustive sampling of each locus for each taxon in a given phylogenetic study, and especially so when considering supermatrices. The power behind these analyses lies in their ability to tie together disparate but already phylogenetically information-rich datasets, and the generation of fully sampled alignments is not required to do this successfully (Wiens 2003). This is simply because, given the sampling patterns already present in the data, only a small subset of the potential additional sequence data will meaningfully impact the resulting topology. For instance, it is not advisable to exhaustively sample slowly evolving markers for closely related taxa because a lack of signal makes many of these sequences almost completely redundant, and similarly, it makes little sense to sequence rapidly evolving markers for distantly related taxa, because very high levels of random signal in these data can be expected to limit their phylogenetic utility. It is worth pointing out however that despite the power of available methods to accurately reconstruct phylogenetic relationships in the face of high levels of missing data, these sparse alignments can negatively impact our ability to accurately infer branch lengths (Lemmon et al. 2009, but see Wiens and Morrill 2011), and this could present problems for studies intending to generate ultrametric trees.

While methodically filling an entire supermatrix alignment may not prove the most efficient use of resources, careful addition of sequence data to sparse alignments may dramatically improve results (van der Linde et al. 2010). Adding select loci for underrepresented taxa may be one way to improve not only the inclusivity of the resulting trees, but also the decisiveness and phylogenetic informativeness (Townsend and Leuenberger 2011; López-Giráldez and Townsend 2011; Townsend and López-Giráldez 2010) of the alignment itself. Several recent studies have addressed questions regarding sampling optimization strategies for phylogenetic inference (Townsend and López-Giráldez 2010; Sanderson et al. 2010; Thomson and Shaffer 2010; Yan et al. 2005; Cho et al. in press), and additional inquiry into this
field will no doubt yield practical information for frugal systematists wishing to improve the utility of their datasets while conserving valuable materials and time.

**Systematics of Cyperaceae**—Our results corroborate many previous findings regarding sedge phylogeny, but provide new and important insights that contradict some earlier studies. One of the largest differences between this and the next most-inclusive sedge phylogeny in the literature (Muasya et al. 2009) is that the tree we present here has very strong support for nearly all relationships among major clades. Several differences in topology also exist. Classification units referred to herein follow Simpson et al. (2007) and Goetghebeur (1998). For a thorough, concise discussion of Cyperaceae classification, refer to Muasya et al. (2009) and for more detailed taxonomy, Govaerts et al. (2007).

In the best-resolved sedge phylogeny we present here (Figs. 3-6), subfamily Mapanioideae is strongly supported as sister to the rest of the Cyperaceae, with the Trilepideae sister to the all remaining, just as in Muasya et al. (2009) (Fig. 3). However, we find a Sclerieae + Bisboeckelereae clade strongly supported as sister to all remaining Cyperaceae, which matches expectations based on morphological classifications, but is incongruous with Muasya et al. (2009), where these tribes were nested within the Schoeneae. Like Muasya et al. (2009) however, we do resolve Didymiandrum and Lagenocarpus, members of tribe Cryptangieae that have previously been classified with Scleria (Goetghebeur 1998), to be nested within the Schoeneae with very high support (Fig 3). We find support for this Schoeneae + Cryptangieae clade as sister to all other remaining lineages, and strong support for the inclusion of the genera *Cladium* and *Gymnoschoenus* within the Schoeneae clade itself (Fig. 3), again in contrast to Muasya et al. (2009), where these genera formed a grade below the remaining Schoeneae. Although relationships within the Schoeneae are generally unresolved, several well-supported clades corresponding to major genera are present, though some genera (e.g. Gahnia, Tetraria) are clearly polyphyletic. There is good support for a sister relationship between *Cladium* and the rest of the tribe (Fig. 3). The genus *Rhynchospora* is strongly supported as sister to all remaining lineages, and contains all present members of the genus *Pleurostachys* (Fig. 3).
Strong support exists for one large additional major clade consisting of two sister subclades: 1) Cariceae + Dulichieae + *Khaosokia* + Scirpeae; and 2) Abildgaardieae + Cypereae + *Eleocharis* + Fuireneae (Figs. 4-6). Within the first of these, Dulichieae is supported in a sister relationship with a Cariceae + *Khaosokia* + Scirpeae clade, but relationships between Cariceae, Scirpeae, and the monotypic *Khaosokia caricoides* are unresolved. The monophyly of the genus *Carex*, if circumscribed to include *Cymophyllus*, *Kobresia*, *Schoenoxiphium*, and *Uncinia* (i.e. the Cariceae), is strongly supported (Figs. 4 and 5). Most species of Scirpeae present in this dataset are not represented by sequences at rapidly-evolving, data rich markers such as *ndhF* and *ITS*, and the addition of these markers may help clarify the relationships within this group and among Cariceae, *Khaosokia*, and Scirpeae. The monophyly of a clade containing the Abildgaardieae and *Eleocharis* is well supported, as is the monophyly of the aforementioned subclades within it (Fig. 5). Relationships among the genera *Bolboschoenus*, *Fuirena*, *Schoenoplectiella*, and *Schoenoplectus* are poorly resolved (Figs. 5 and 6), and additional studies involving novel sequencing effort will likely be necessary to shed light on this problematic area. Within the Cypereae, a strongly supported sister relationship is indicated between a *Ficinia* + *Isolepis* clade in which these two genera form nearly monophyletic groups, and a clade in which *Scirpoides* is sister to a clade containing a highly paraphyletic *Cyperus* with numerous other genera interdigitated throughout it (Fig. 6).

Relationships within major clades are in general less well-resolved than those in the studies from which these data were originally published (Hinchliff et al. 2010; Roalson et al. 2010; Hinchliff and Roalson 2009; Muasya et al. 2009; Starr and Ford 2009; Thomas et al. 2009; Waterway et al. 2009; Ghamkhar et al. 2007; Katsuyama et al. 2007; Simpson et al. 2007; Chacón et al. 2006; Verboom 2006; Yano et al. 2004; Zhang et al. 2004; Muasya et al. 2002; 2001; Roalson et al. 2001; Roalson and Friar 2000; Yen and Olmstead 2000). This may potentially be due to decreased levels of matrix decisiveness for shallow branches because of the addition of species that lack overlap with the genes sampled for the more detailed original studies. For instance, a single exemplar species *S* of genus *A* may be used to root a detailed study of some other genus, for which genes *W* and *X* are sampled. Combining the data from this study with those from a study of genus *A* containing only data for genes *Y* and *Z* creates a scenario where
the placement of $S$ within $A$ cannot be determined because no homologous sites exist on which to base comparisons of $S$ to any other tip in $A$. Issues of data non-overlap such as this become complex when combining multiple datasets, and may lead to the collapse of clades for which strong supporting signal is otherwise present in the alignment (Sanderson et al. 2010). One additional factor likely to contribute to low resolution near the tips of the tree is the decrease in the amount of phylogenetic information available to inform branches at shallower depths, because of the smaller numbers of taxa (and therefore also fewer sampled loci) that can be used to resolve these nodes.


Figure 1. Heat map of nucleotide sequence sampling in Cyperaceae in GenBank release 185. Sampling density and summary statistics for all currently recognized genera of Cyperaceae (x-axis) are shown for each of the twenty-three most densely sampled genetic markers (y-axis) on GenBank. Sequence data were gathered from GenBank and aligned using phlawd. Sampling density for each marker/genus combination is a proportion of the total number of species sampled for that marker/genus divided by the total number of species currently recognized in that genus. Values range from 0 (no species sampled) to 1 (all species sampled). Summary statistics are used to rank to relative quality of taxonomic sampling across the family. “TOT percent cover” is the proportion of total species sequences obtained for some genus. “TOT proportion spp” is the relative contribution of a genus to the species richness of the family. “ABS importance” is a summary statistic that incorporates both TOT percent cover and TOT proportion spp to estimate the value of additional sequencing within some genus in order to improve taxonomic coverage for the family itself. “RANK importance” is the percentile rank of each value of ABS importance, which represents the relative necessity of additional sequencing investment across all genera. All these values range between 0 (low) and 1 (high). See Appendix I for a more thorough explanation including formulas used to calculate summary statistics. Tree diagrams are distance networks representing the similarity of sampling patterns among genera and loci, not phylogenetic relationships.
Figure 2. Majority-rule consensus trees resulting from parallel 300-replicate ML bootstrap searches performed on each of the alignments summarized in Table 2. Letters A-D correspond to rows 1-4 from Table 1, respectively. Although tip and branch labels could not be included due to space constraints, several topological landmark nodes corresponding to several major clades are labeled. The star labels the genus *Carex*, the square *Eleocharis*, the circle the tribe Cypereae, and the triangle the tribe Schoeneae. The dotted lines in C and D indicate an area of the tree that apparently experienced a decrease in
resolution as a result of more stringent rogue taxon filtering, although as discussed in the text, this apparent loss of resolution may actually represent an increase in phylogenetic accuracy.
Figure 3. Caption refers to Figures 3-6. Majority rule consensus tree resulting from a 300-replicate ML search performed on the alignment corresponding to Table 1, row 4. Branch labels indicate bootstrap.
proportions. Major clades and grades of the Cyperaceae are labeled to the right. Abbreviations are as follows: D. = Dulichieae; Fuir. grade = Fuireneae grade; T. = Trilepideae. Each labeled group occurs only once in the tree, but it may extend across multiple pages. In these cases, the parts of the group on different pages are labeled individually.
Figure 4. Refer to Fig. 3 caption.
Figure 5. Refer to Fig. 3 caption.
Figure 6. Refer to Fig. 3 caption.
APPENDIX A: LIST OF VOUCHERED DNA SAMPLES FOR CHAPTER 1

Key: [Species] (Subgenus): [Id number for this study] / [Collector] [Collector’s number]. [Year].
  [Country], [Herbarium], [ITS GenBank number], [trnC-ycf6 GenBank number], [ycf6-psbM
  GenBank number].

_E. acutangula_ (Roxb.) Schult. subsp. _acutangula_ (_Limnochloa_): 126 / P. M. Peterson, C. R. Annable

_E. acutangula_ (Roxb.) Schult. subsp. _breviseta_ D.J. Rosen (_Limnochloa_): 376 / J. Elgaard Madsen 5455.

_E. acutangula_ (Roxb.) Schult. cf. subsp. _neotropica_ D.J. Rosen (_Limnochloa_): 377 / M. Rimachi Y.

_E. brassii_ S.T. Blake (_Limnochloa_): 226 / C. E. Hinchliff 50. 2006. Australia. WS. FJ826566, FJ829373,
FJ829416.

FJ826596, FJ829403, FJ829446. ; 220 / C. E. Hinchliff 22. 2006. Australia. WS. FJ826599,
FJ829406, FJ829449. ; 231 / C. E. Hinchliff 88. 2006. Australia. WS. FJ826597, FJ829404,

_E. elongata_ Chapman (_Limnochloa_): 76 / E. H. Roalson 1498. 2002. USA. WS. FJ826592, FJ829399,
FJ829442.

_E. equisetoides_ (Elliot) Torr. (_Limnochloa_): 137 / C. T. Bryson 16554. 1998. USA. MO. FJ826585,
FJ829392, FJ829435.

FJ826582, FJ829389, FJ829432. ; 205 / E. H. Roalson 1557. 2006. Mexico. WS. FJ826583,
FJ829390, FJ829433. ; 208 / E. H. Roalson 1560. 2006. Mexico. WS. FJ826581, FJ829388,
FJ829431.


APPENDIX B: LIST OF VOUCHERED DNA SAMPLES FOR CHAPTER 2

Key: [Species][Authority]: [Id number for this study] / [Country], [Year]. [Collector] [Collector’s number], [Herbarium]. [ndhF GenBank number], [psbB- psbH GenBank number].


**Calyptrocarya** sp.: 486 / Guyana, Potaro-Siparuni, 2006. K. M. Redden s.n., US. GU075473, GU075577.


**Caustis dioica** R. Br.: 339 / Australia. M. W. Chase 2225, K. GU075499, GU075603.

**Cladium mariscoides** (Muhlenberg) Torrey: 342 / Brasil. Thomas & al. 10403, K. GU075503, GU075607.


**Cyperus kerstenii** Boeck.: 328 / Kenya. M. Muasya 984, K. GU075455, GU075559.

**Cyperus longus** L.: 323 / Australia. M. W. Chase 2276, K. GU075452, GU075556.
**Cyperus sphacelatus** Rottb.: 301 / Guyana, Potaro-Siparuni, 2006. K. M. Redden 3886, US. GU075451, GU075555.


**Diplacrum capitatum** (Willd.) Boeck.: 489 / Guyana, Potaro-Siparuni, 2006. K. M. Redden 4033, US. GU075475, GU075579.


**Eleocharis cylindrostachys** Boeck.: 233 / Australia, Queensland, 2006. C. E. Hinchliff 93, WS. GU075420, GU075524.

**Eleocharis dulcis** (Burm. f.) Trin. ex Hensch.: 231 / Australia, Northern Territory, 2006. C. E. Hinchliff 88, WS. GU075440, GU075544.

**Eleocharis elegans** (Kunth) Roem. & Schult.: 204 / Mexico, 2006. E. H. Roalson 1556, WS. GU075425, GU075529.


Eleocharis pallens S.T. Blake: 228 / Australia, Northern Territory, 2006. C. E. Hinchliff 76, WS. GU075419, GU075523.


Ficinia nodosa (Rotth.) R. Br.: 500 / L. Macleod 1, NY. GU075458, GU075562.


**Fimbristylis** cf. **pauciflora** R. Br.: 225 / Australia, Northern Territory, 2006. C. E. Hinchliff 40, WS. GU075446, GU075550.

**Fimbristylis** sp.: 282 / Australia, Northern Territory, 2007. C. E. Hinchliff 4, WS. GU075441, GU075545.


**Fuirena** sp.: 333 / Brasil? Thomas 10404, K. GU075466, GU075570.


Mapania macrophylla (Boeck.) H. Pfeiff.: 288 / Guyana, Potaro-Siparuni, 2006. K. J. Wurdack 4213, US.
GU075497, GU075601. 305 / Guyana, Potaro-Siparuni, 2006. K. M. Redden 3988, US.
GU075496, GU075600. 492 / Guyana, Potaro-Siparuni, 2006. K. M. Redden 4065, US.
GU075498, GU075602.


Pycreus nuerensis (Boeck.) S.S. Hooper: 327 / Tanzania. M. Muasya 940, WS. GU075454, GU075558.

GU075488, GU075592.

GU075490, GU075594.


Rhynchospora cf. corymbosa (L.) Britton: 475 / Bolivia, Beni, 2007. C. E. Hinchliff 656, WS.
GU075489, GU075593.

GU075482, GU075586.


Scleria sp.: 307 / Guyana, Potaro-Siparuni, 2006. K. M. Redden 4000, US. GU075505, GU075609. 487 /


Uncinia hamata (Sw.) Urb.: 472 / Bolivia, La Paz, 2007. C. E. Hinchliff 527, WS. GU075479, GU075583.

GU075435, GU075539.

Zameioscirpus muticus Dhooge & Goetghebeur: 474 / Bolivia, La Paz, 2007. C. E. Hinchliff 648, WS.
GU075516, GU075620.
Equations for sampling summary statistics

C. E. Hinchliff

November 7, 2011

The percent coverage of a given genus $k$ (TOT percent cover):

$$c_k = \frac{\sum_{i=1}^{m} n_i}{m \cdot n_k}$$  \hspace{1cm} (1)

where $m$ is all markers in the study, $n_i$ = the number of species sequenced for marker $i$, and $n_k$ = the number of species in the current genus.

The total proportion of species in a given genus $k$ (TOT proportion) is simply

$$d_k = \frac{n_k}{\sum_{g=1}^{g} n_k}$$  \hspace{1cm} (2)

where $g$ is all genera.

The absolute importance value assigned to a genus is a summary statistic that incorporates the current sampling coverage and the taxonomic contribution of each genus to total species richness. It is calculated for a genus $k$ as:

$$M_k = 1 - (c_k \cdot (1 - d_k))$$  \hspace{1cm} (3)

$M_k$ increases toward 1 for genera representing a large proportion of the species richness of the family and for those that have been poorly sampled, and decreases toward $d_k$ for those that have been sampled exhaustively. A value of 1 (the highest importance) is assigned for any genus with zero percent coverage ($c_k = 0$), regardless of its contribution to the total species richness of the family.